Using Attentional Highlighting to Guide Visual Categorization

Carlos Díaz Hernández

Director: Michael C. Mozer (University of Colorado Boulder)
Ponent: Javier Béjar Alonso (Universitat Politècnica de Catalunya)

June 18 2015

ABSTRACT

Many visual categorization tasks require domain-specific expertise. Our objective is to aid novices in making these difficult categorizations. We explore the usefulness of guiding individuals to diagnostic regions via the use of spatial filtering, what we refer to as attentional highlighting.

Diagnostic regions for each image were derived from a dataset provided by Deng, Krause and Fei-Fei (2013). Subjects made a binary classification decision of a test image by clicking on one of two reference images. One of reference images was the same category as the test image; the other was an incorrect lure. Subjects were randomly assigned to one of two conditions that involved either unaltered or highlighted reference images. All subjects made classification decisions on four trial types. Trial types ranged from highly similar (difficult) categories to highly dissimilar (easy) categories.

We compared accuracy and response time across the between-subject conditions and across within-subject trial types. Accuracy did not vary across the two conditions. Attentional highlighting decreased response times by approximately 750 ms for the hardest trial type. These results suggest that attentional highlighting can be used to reduce the temporal costs associated with making difficult visual category distinctions.

RESUMEN

Muchas tareas de categorización visual requieren conocimientos de dominio específico. Nuestro objetivo es proporcionar ayuda a personas no expertas a la hora de realizar estas difíciles categorizaciones. Exploramos la utilidad de guiar a las personas hacia regiones diagnósticas de imágenes a través de la utilización de filtrado espacial, a lo que nos referimos como *attentional highlighting*.

Las regiones diagnósticas de las imágenes se extrajeron de un conjunto de datos proporcionado por Deng, Krause y Fei-Fei (2013). Los sujetos del estudio realizaron una categorización binaria de una imagen seleccionando una de las dos imágenes de referencia. Una de las imágenes de referencia era de la misma categoría que la imagen de la prueba; y la otra pertenecía una categoría diferente. Los sujetos fueron asignados aleatoriamente a una de las dos condiciones: imágenes de referencia alteradas o no alteradas. Todos los sujetos clasificaron imágenes en cuatro tipos de prueba. Cada tipo de prueba tenía una dificultad diferente: de muy difícil (categorías muy similares) a muy fáciles (categorías muy diferentes).

Comparamos la precisión y el tiempo de respuesta entre todas las condiciones intra-sujeto e inter-sujeto del experimento. Encontramos que la precisión no varió a través de las dos condiciones inter-sujeto. Sin embargo nuestra técnica de *attentional highlighting* redujo los tiempos de respuesta en las pruebas más difíciles en aproximadamente 750 ms. Estos resultados sugieren que el resaltado

atencional puede utilizarse para reducir costes temporales a la hora de realizar categorizaciones visuales difíciles.

RESUM

Moltes tasques de categorització visual requereixen coneixements de domini específic. El nostre objectiu és proporcionar ajuda a persones no expertes a l'hora de realitzar aquestes difícils categoritzacions. Explorem la utilitat de guiar les persones cap a regions diagnòstiques d'imatges a través de la utilització de filtrat espacial, al que ens referim com *attentional highlighting*.

Les regions diagnòstica de les imatges es van extreure d'un conjunt de dades proporcionat per Deng, Krause i Fei-Fei (2013). Els subjectes de l'estudi van realitzar una categorització binària d'una imatge seleccionant una de les dues imatges de referència. Una de les imatges de referència pertanyia a la mateixa categoria que la imatge de la prova; i l'altra pertanyia una categoria diferent. Els subjectes van ser assignats aleatòriament a una de les dues condicions: imatges de referència alterades o no alterades. Tots els subjectes van classificar imatges en quatre tipus de prova. Cada tipus de prova tenia una dificultat diferent: de molt difícil (categories molt similars) a molt fàcils (categories molt diferents).

Vam comparar la precisió i el temps de resposta entre totes les condicions intra-subjecte i inter-subjecte de l'experiment. Trobem que la precisió no va variar a través de les dues condicions inter-subjecte. No obstant això la nostra tècnica de attentional highlighting va reduir els temps de resposta en les proves més difícils en aproximadament 750 ms. Aquests resultats suggereixen que el ressaltat atencional pot utilitzar-se per reduir costos temporals a l'hora de realitzar categoritzacions visuals difícils.

"Look up at the stars and not down at your feet. Try to make sense of what you see, and wonder about what makes the universe exist. Be curious."

— Stephen Hawking

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Prof. Michael C. Mozer, whose expertise, understanding, and patience, added considerably to my research experience.

A very special thanks goes out to my colleague Brett D. Roads, without whose motivation, encouragement and guidance I would not have been able to complete this endeavor.

I must also acknowledge Mr. Pete Balsells and the University of Colorado Boulder for going that extra mile to help students from Catalan universities come to the United States to conduct research projects.

I would also like to thank my family and friends for their support and encouragement across the ocean.

TABLE OF CONTENTS

FIGURES

CHAPTER I. INTRODUCTION

The inclusion of support systems in applications is becoming more and more common in training and professional tools. Mobile devices like smartphones or tablets have become very popular, and that has lead to the booming of applications that provide solutions to any kind of problem that one can imagine. As it could be expected, there are applications that provide help to people in the context of a decision making process.

In this context, we aim to aid non-experts in making visual categorizations that require domain-specific expertise.

We explore the usefulness of guiding individuals to diagnostic regions via the use of image manipulations. In doing so, we expect individual to learn how to direct their attention to these regions thus enhancing the learning process.

This document describes the entire process of this project, as well as its outcome and implications. This introductory chapter analyzes the objectives and the context of the project. The second chapter is about the managerial aspects of the project. The third chapter describes how the plan was executed. Finally the conclusion chapter provides a summary of the achievements and their implications.

1. Objectives

The main problem we are trying to solve is to determine whether presenting images with certain image manipulations (attentional highlighting) has a positive effect on performance when a person is asked to categorize visual stimuli.

What does improving performance mean? Better performance means more accuracy, and quicker response times. Hence, our main problem develops into two different objectives: **maximizing correctness** and **minimizing response time**s.

To do this we will need to build a feature highlighting system. This system will guide individuals to discriminating regions by manipulating the visual stimuli. We need to be able to direct the attention to those diagnostic features that will help humans categorize images correctly and in a short time.

A broader goal of the project is to be able to use this improvement in performance to apply our system to learning tasks. One application could be, for example, a software for learning about plant families, or and aid system for the visually impaired.

2. Scope

This section describes what the project involves and what it does not. For the purpose of achieving the abovementioned objectives, some parts of the scope are subject to change depending on preliminary results.

In order to be able to solve a problem of the nature of the one we are proposing, we first need to define which image manipulations will affect visual perception in a positive way and be helpful to our task. In order to do this we will need to consult the scientific literature on visual perception, cognitive science and decision-making processes.

Our system must be able to detect diagnostic regions of the visual cues. This means to determine what parts of the image will help subjects place the images in the right category more easily. This requires an in-depth analysis of our visual cues and the use of the results of previous studies.

This project also involves designing and running experiments on human subjects in order to prove our hypotheses. This includes defining what the conditions we want to test are as well as determining a behavioral measure so that we can compare the outcomes of the different conditions. The outcome of these experiments must provide reliable data that will be analyzed in order to conclude and extract implications regarding our hypothesis.

This project does not include using a software tool for the end user, but to provide a system that can be applied in such tools. However, depending on time and resource constraints the project director may decide to build a test application with a more specific end-user application.

3 Context

This section contextualizes the project by providing a description of the relevant areas of interest and explaining which are the actors affected by the development of this project.

3.1 Areas of interest

There are several scientific fields that are relevant to the contextualization of the project. Specifically, these are Human decision processes and Visual perception,

3.1.1 Human decision processes

In this project, we are trying to modify images in a way that it affects a human decision process, more specifically, categorizing items using a visual stimulus. In order to be able to develop our project we need to know how these decision processes work.

We can consider decision-making to be a cognitive process that results in selecting a course of action among other alternative possibilities. Every decision-making process produces a final choice that may or may not prompt action.

Decision-making is the study of identifying and choosing alternatives based on the values and preferences of the decision maker [1].

3.1.2 Visual perception

Visual perception can be defined as the ability to interpret the surrounding environment by processing information that is contained in visible light. The resulting perception is also known as eyesight or vision [2].

This field is important to us since we need to know how visual perception works in order to be able to manipulate images to highlight certain features of an item.

Regarding visual perception, we are specifically interested in object recognition via visual stimuli. The bottom-up theory for visual perception suggests that objects are recognized using some sort of salience map that prioritizes information of key features in an object. This is what we will try to simulate when manipulating our images.

4 State of the art

This section is to provide an overview of the state of the art that is relevant to this project.

4.1 Decision support systems

Decision support systems are computer-based information systems that provide help in the process of making decisions. The nature of these decisions comprises a wide range of types from business decisions to medical decisions. Our project is aimed at improving decision latencies and accuracy when presented a visual stimulus.

Berg et al. describe a decision support system that point out visual diagnostic regions on birds. In a task such as bird species recognition, automatic recognition systems can now exceed the performance of non-experts - most people are challenged to name a bird species, let alone identify them. [3] This leads us to the question, "Can a recognition system show humans what to look for when identifying classes?"

Another relevant state-of-the-art tool regarding this topic is the Leafsnap mobile app. This app is a visual recognition system for automatic plant species identification. The system helps users identify trees from photographs of their leaves. As described by Kumar et al., Leafsnap was developed to greatly speed up the manual process of plant species identification, collection, and monitoring [4].

4.2 Data collection methods

A/B testing is a traditional method of conducting a randomized controlled experiment to compare the effect of two treatments, A and B, on human subjects. For example, two alternative banner ads may be served to evaluate which is more effective in driving click through. A/B testing is the dominant paradigm in the experimental behavioral sciences used to understand human learning, reasoning and decision-making.

However, even though this method can be extended to compare a many different situations, it does not solve a very common problem: searching over a large, possibly combinatorial or continuous space of alternatives to identify the treatment that achieves the best outcome. One solution to this problem is using Gaussian process.

According to the textbook *Gaussian Processes in Machine Learning* by Rasmussen et al., this process views a function as a very long vector, each entry in the vector specifying the function value f(x) at a certain input x. If you test the outcome for a finite number of values of x, then inference in the Gaussian process will provide a fairly accurate function values for the rest of the continuum [5].

4.3 Image manipulations

Image manipulation techniques have been around for many years now. We plan to include both spatial frequency and saturation modifications in our images. There are several state-of-the-art theories and experiments that justify our decision to use these two parameters (level of spatial frequency manipulation and level of saturation manipulation) to try and find the optimal manipulation of an image to help people make the right categorizing decision as fast as possible.

Firstly, a study by Itti and Koch justifies the use of image manipulations as a whole. Their study presents a state of the art theoretical framework for thinking about visual attention. This theoretical framework supports the idea that image manipulations should help direct people's attention [6]. This justifies our idea that modifying certain aspects of a visual stimulus can affect response time.

Secondly, to justify the first parameter I have chosen to present Zhang's study [7] about saliency. The saliency of an item is the state by which it stands out to its relative neighbors. This study proposes a system for implementing a system that computes visual saliency based on the framework described by [6]. They define saliency by considering what the visual system is trying to optimize when directing attention [7]. This justifies our idea of modifying spatial frequency to blur parts of the images that are not key features that humans use to categorize objects.

Thirdly, In order to justify the parameter of saturation manipulation, we have found literature on the results of the experiment described in *The role of color in expert object recognition* [8]. In this experiment bird images were shown in three different conditions: natural color, non-natural color, and gray scale. Their results show that the color affected the performance of bird experts and bird novices when recognizing birds at their family level[8].

4.4 Visual categorization

Our main problem is indeed a visual categorization problem. Visual categorization is the ability to classify visual stimuli into discrete functional groups and is an essential component of recognition and decision making.

A recent review of visual category learning provides support for the idea that learning the relevant features is important for visual categorization. Therefore, guiding individuals to diagnostic features may help them perform the task. [9]

CHAPTER II. MANAGEMENT ASPECTS OF THE PROJECT

This chapter is dedicated to the managerial aspects of the project. It contains information on the methodology, validation process, temporal planning, budget, sustainability, and legal aspects.

1 Methodology

This section describes the methodology according to which we intend to develop this project so that we are able to achieve all the objectives we have set.

This project has been born in an academic environment. This is why it will be guided by empirical data that will be gathered in a series of experiment. The specifics of these experiments are to be decided in our weekly lab meetings. The following points were agreed on at the initial stages:

1.1 Baseline experiment

We need to set up a baseline condition in order to be able to measure the improvement and success of our project. The way we will set up this baseline is by running an experiment on a reasonable pool of subjects. These subjects will be given series two sets of images and one test image. The two sets of images correspond to two different kinds of birds, respectively. The test image correspond to one these two kinds of birds. The test image has not been manipulated. In each trial, the subjects will have to decide whether the bird on the test image correspond to the category of the bird in one set of images or the other.

The outcome of this experiment will set the bar we need to beat using our optimal manipulation scheme. The results must include accuracy and response time per trial.

1.2 Definition of image manipulations

We need to determine what kind of manipulations we want to apply to our images if we want to improve performance. For this, we will have to research how visual perception works.

There are many different manipulations that we could apply. Examples of these image manipulations are: modifying the saturation, blurring out, circling or coloring certain parts of the image.

Literature seems to suggest that the visual process is a bottom-up procedure. This is way we need to use these manipulations to create some sort of perception of saliency giving more priority to diagnostic parts of the image. Which leads to the following subsection that explains how we plan to determine diagnostic parts.

1.3 Determining diagnostic parts

It is very important that we highlight the right parts of the image so that we are not misleading subjects attention to parts of the image that are not meaningful for the purpose of discriminating categories.

We will be using the data set of bird images collected when testing the “Bubbles” games. In this game, subjects were presented a setting similar to our baseline experiment. However, the test image was completely blurred. The subjects taking part in this experiment had to select the parts of the test image they wanted to “unblur” before they made a categorization decision.

This dataset contains enough information for us to build a heat map of the regions users thought were more diagnostic.

1.4 Final experiment

Having established a performance goal, having decided on image manipulations and having detected diagnostic regions for the images in our dataset. We will run a multiple-condition experiment testing different levels of image manipulations and how they affect subject performance.

1.5 Further experimentation

Depending on the results of our experiments, the team may agree on try different manipulations or to run more experiments in order to further validate the outcome.

2 Validation

During the early stages of the development of this project, we seek expert validation of our decisions. We will also be making informed decisions based on the existing scientific literature.

The effectiveness of our strategies will be tested out on human subjects. The results will be then statistically analyzed.

3 Development tools

There are a number of development tools at our disposal. The following are the ones we have selected to carry out this project. Here we expose the reasons why we have chosen to used these tools and not others.

3.1 Matlab

Matlab and Simulink are the tools that we will be using for the analysis of existing data as well as for the generation of image manipulation (Using Matlab's Image Processing Toolbox). Matlab has been chosen for being known to all members of the team as well as for its endless analysis capabilities. The University of Colorado Boulder provides free Matlab and Simulink licenses for academic purposes.

3.2 PsiTurk

The experiments will be programmed in JavaScript and adapted to PsiTurk for their use in the Amazon Mechanical Turks platform .

There are several ways to conduct behavioral experiments. The easiest option is probably to design your own program and to gather people into a room and have them take the experiment. Payment can be done instantly, your own program allows you to collect any information you need and you can always be sure that subjects are only paying attention to the experiment. However, this requires a lot more money and time as well as a higher difficulty to find subjects.

The second option is to run experiments over the Internet. The advantages of this option are many as well. For starters, you do not need a physical facility where subjects take your experiments. Also, time is not a constraint since subjects can take the experiment at their desired time. The disadvantages are that subjects perform the experiments in a less controlled environment and that if you want guarantees you need to use an existing platform like Amazon Mechanical Turk.

Implementing experiments for Amazon Mechanical Turk is a fairly difficult task. Nevertheless, we opted to use psiTurk for Amazon Mechanical Turk. PsiTurk acts as an intermediate step between your JavaScript code and the requirements of Amazon Mechanical Turk. We chose to run our experiments over the Internet because it is a common and safe practice nowadays that produces reliable results and although it requires a bit more of learning, it ends up saving time and money in the end.

4 Possible obstacles and risks

The main obstacle we could find is that despite that the existing scientific literature backs up our initial hypotheses, the results of the experiment may determine that our system does not produce significant improvement in performance over non-manipulated images. This is very unlikely but should this happen we would have to come up with new strategies. However, the future progress of this experiment will ultimately be based on the team leader's discretion.

Noisy data could also pose a problem. Since we are running our experiments through Amazon Mechanical Turk, we need to control parameters such as latencies in response times and defocusing of the window in order to detect cheating.

On a more specific level, we might find that our performance results do not adjust well to a certain distribution so our Gaussian Process Regression will not produce a very accurate outcome. We may have to try different models or change or experimentation scheme to try and find decent solution using the traditional scientific experimentation scheme.

5 Temporal Planning

The aim of this section is to define and describe the tasks pertaining to this project as well as to establish the time allocation, associated resources and scheduling of them. In the last section of this document there is an explanation of how we plan to execute this temporal planning so that we finish our project before our deadline.

The starting date of the project is January 20th, 2015 and the deadline is May 31st, 2015.

5.1 Description of tasks

This section lists and describes the series of tasks that conform project.

5.1.1 Initial research and feasibility

This is an initial task that is required for every research project. The aim of this task is to gather and analyze existing information that is pertinent to our project, such as scientific articles on visual perception, or previous studies on decision-making processes, etc.

At this initial stage this task involves brainstorming sessions in order to come up with possible ways to carry out the project as well as to assess whether it is feasible to do this in the time given.

This task will be identified as T1.

5.1.2 Planning of the project

Once we have established an agreement on the basic terms of the project, we need to clearly state our objectives and scope of the project. The objectives and scope must be set according to the time and resources available. The planning involves allocating time and resources to tasks and will serve as a guideline throughout the development of the project.

This task will be identified as T2.

4.1.3 Analysis of existing data

At this stage of the project we will assess existing data from previous studies to determine what is valuable to our cause. This includes, determining whether we want to use an existing dataset for our experiments or if it is worth it to generate our own.

This task will be identified as T3.

4.1.4 Baseline experiment

This tasks involves the design, implementation and deployment of the baseline experiment

This task will be identified as T4 and it is divided as follows:

Design of experiment

Taking into account what was agreed on during the first stages of the project. The aim of this experiment is to collect data on accuracy and response times under a control condition. This means that the images will be presented with no manipulations.

The user interface needs also needs to be agreed on during this task.

JavaScript adaptation of baseline experiment

For the purpose of being able to run the baseline experiment on Mechanical Turk (PsiTurk), we will need to implement it in JavaScript.

After implementation we need to run the experiment on our local server to catch any errors and solve them before deployment. Finally, when everything is ready we will deploy the experiment on Mechanical Turk.

Recollection of results

We must collect exhaustive data and metadata of the experiment. This will allow us to rule out subjects that were not paying attention to the task, and to have more significant results to compare with the future experiment results: response latencies, accuracy, out of window times, etc.

4.1.5 Defining image manipulations

After consulting psychologists and doing our research we will define which image manipulations we deem best for achieving our objectives.

Taking into account budgeting and time factors we will decide how many conditions we are willing to test.

This task will be identified as T5.

4.1.6 Level of manipulation experiments

This task describes how manipulation experiments will be conceived. Although it is considered a single task in this document, more than one condition will be tested. This description applies to all those experiments.

This task will be identified as T6 and it is divided as follows:

Design of experiment

For every image manipulation we want to test different levels of manipulations, we need to decide on this scale so that we can later make an accurate inference of the optimal parameters.

JavaScript adaptation of the experiments

For the purpose of being able to run the experiments on Mechanical Turk, we will need to implement it in JavaScript.

After implementation we need to run the experiments on our local server to catch any errors and solve them before deployment. Finally, when everything is ready we will deploy the experiments on Mechanical Turk

Recollection of results

We will need to collect from these experiments least the same information we gathered from the baseline experiment so that we are able to perform an appropriate analysis of the results

4.1.7 Analysis of results and reassessment

After collecting the results from each different condition we want to use Gaussian Process Regression to infer the optimal level of manipulation for each manipulation applied to the images. It involves trying different data distributions and its outcome will define the condition for the last experiment.

This task will be identified as T7.

4.1.8 Final experiment. (Validation of results)

We might need to run a final experiment to replicate and further validate the results obtained in the previous tasks

As with the other experiments we will need to design, implement, test and deploy on Mechanical Turk.

After we have collected the results, we will need to determine whether we consider them satisfactory.

This task will be identified as T8.

4.1.9 Weekly meetings

Every Tuesday from 2pm to 3:30pm we will allow time for a weekly meeting to report on progress and errors and, when necessary, determine mitigation strategies.

This task will be identified as T9.

4.2 Time table

Now that all the tasks have been described, we allocate time (in hours) to each of the tasks in the section above. The following table summarizes the time we have considered appropriate for each task.

Task	Time (hours)
T1. Research and feasibility	25
T2. Planning of the project	80
T3. Analysis of existing data	50
T4. Baseline experiment	80
T5. Defining manipulations	60
T6. LOM experiments	180
T7. Inference of optimal parameters	40
T8. Final experiment	60
T9. Weekly meetings	25
Total	600

Table 1 Time per task

4.3 Resources

We have divided the list of resources into hardware and software. Here are both list including the name of the resource and the tasks they are required for.

4.3.1 Hardware

- MacBook Pro 2010. 2,4 GHz Intel Core 2 Duo. NVIDIA GeForce 320M 256 MB. 4 GB 1067 MHz DDR3 (Used in all task)
- Local server. (Used in the testing of experiments in tasks T4, T6 and T8)

4.3.2 Software

- OS X 10.9.5 (all tasks)
- Matlab and Simulink (Tasks T3 to T8)
- Sublime Text 2 (T4, T6 and T8)
- JavaScript (implementation phase of tasks T4, T6 and T8)
- Amazon Mechanical Turk (Deployment phase of tasks T4, T6 and T8)

Additionally, we will need human resources in the form of subjects that will take part in our experiments (tasks T4, T6 and T8).

4.4 Gantt Diagram

This section describes how tasks will be organized in time. First, we will establish dependencies between tasks. Based on these dependencies we will generate a Gantt diagram and establish a timeline for the development of the project.

4.4.1 Task dependencies

The dependencies between tasks are shown in the following graph:

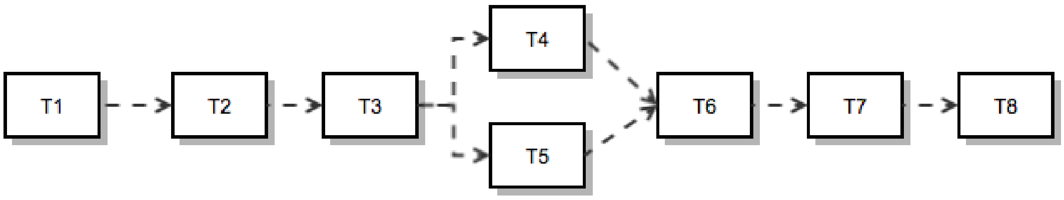


Figure 1 Task dependencies

4.4.2 Task timeline

This table shows the deadlines we have computed for each task

Task	Start Date (MM/DD/YY)	End Date (MM/DD/YY)
T1	03/01/15	03/04/15
T2	03/05/15	03/18/15
T3	03/19/15	03/26/15
T4	03/27/15	04/09/15
T5	03/27/15	04/06/15
T6	04/10/15	05/11/15
T7	05/12/15	05/18/15
T8	05/19/15	05/27/15

Table 2 TImeline of tasks

The following chart shows the timeline according to the time allocation and dependencies in the previous section.

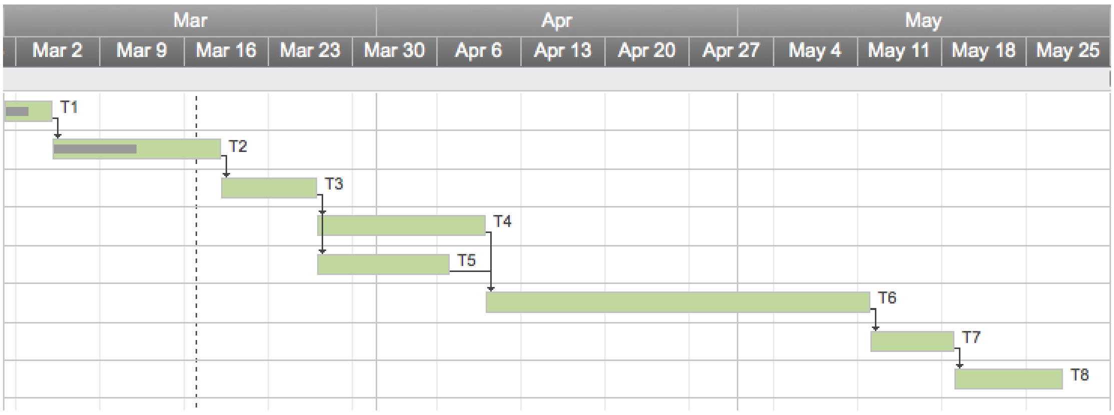


Figure 2 Timeline

4.5 Action Plan

Most tasks will be performed in the order they appear on the Gantt diagram. Our idea is to try and test as many strategies as possible. However, time and budget is limited and we need to use our resources wisely. We will first focus on those manipulations that seem more promising according to experts and existing literature. Given the result and time remaining after these first trials we will decide whether it is necessary to run more experiments. This makes the schedule for *level of manipulation experiments task (T6)* quite flexible. This is also the reason the different phases of this task have not been specified in the diagram.

According to our plan the project would be finished by the end of May. This allows us some time before the hard deadline in mid June to mitigate any obstacle that may arise during the development of the project.

Working at a full time pace of 40 hours per week, the project will take 3 months to complete, according to our calculations.

4.6 Changes in the temporal planning

After defining the image manipulations we were going to apply to obtain attentional highlighting, we received feedback that boosting saturation on diagnostic regions and reducing saturation on non-diagnostic regions was somewhat distracting. During one of the lab meeting sessions we consulted the rest of the members of the lab (who are not involved in the project) and finally decided to drop this manipulation and stick with saliency.

This was a major change in the project since we no longer had the two-dimensional space of image manipulations to which we were going to apply Gaussian optimization. At this point we decided to go in a different direction, which is whether the spatial frequency manipulation would have a different impact, depending on the difficulty of the trial (the more similar two categories are, the more difficult a trial is).

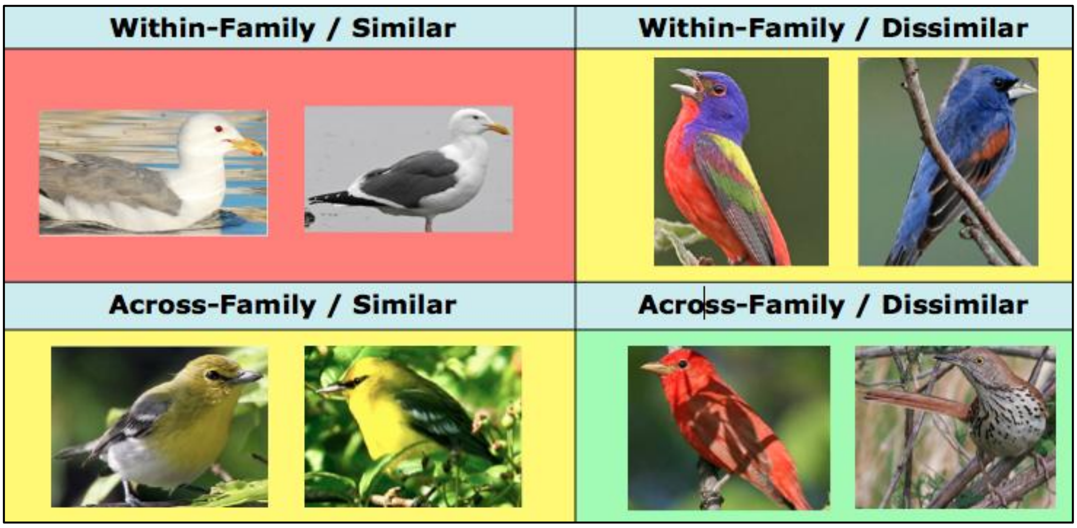


Diagram of trial difficulty based on family similarity and attribute similarity

In this sense we analyzed the attributes of our data set and came up with four levels of difficulty, and agreed on the following experiments.

4.7 New temporal diagram

Even though there have been obstacles concerning the results and we have had to make changes to our planning. We have been able to stick to our deadline. The new timeline looks as follows.

Task	Time (hours)
T1. Research and feasibility	25
T2. Planning of the project	80
T3. Analysis of existing data	50
T4. Baseline experiment	80
T5. Defining manipulations	80

T6. Reassessment of research	20
T7. New Experiments	180
T8. Validation	60
T9. Weekly meetings	25
Total	600

Table 3 Hours per task

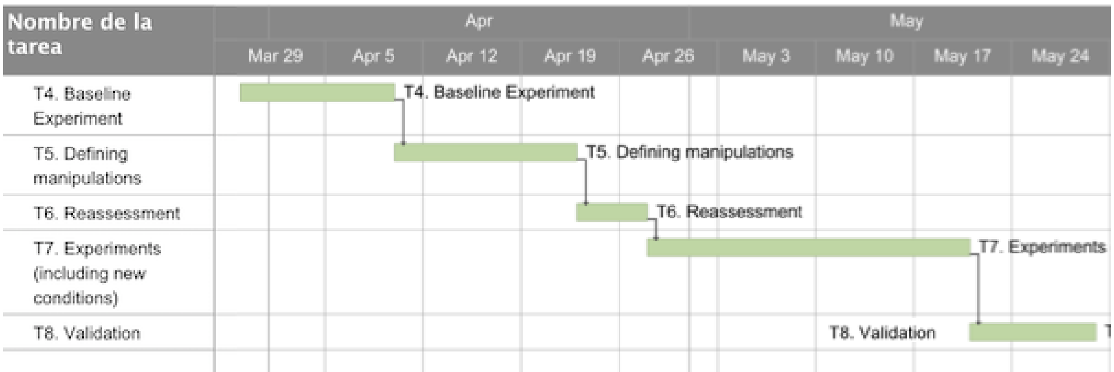


Figure 3 Changed Timeline

4.8 Consequences of changes

The main consequence of the obstacle we had to overcome is that we had to reassess our approach towards our objective. Although we still provide an aid a to a decision support system but we are doing so in a different and more informed way. We now only use one type of image manipulations but we have studied what are the cases in which this aid is more helpful.

The changes on the planning have barely affected the total cost of the project because we have made it possible to finish before our deadline. However, there has been a change in the number of subjects that we have had to pay.

Fortunately, the number of experimental subjects decreases by 20; hence we are indeed spending less money than estimated.

5 Budget

Here we describe the budget that was estimated at the beginning of this project. For this purpose, it contains a section with detailed costs that will be allocated to the different budget categories described in this document. Furthermore, there will be a section explaining how we planned to control the budget and how our contingency budget can be applied.

5.1 Budget estimation

In this section we detail the estimation of costs necessary to be able to bring this project to fruition. The budget is going to be divided in four categories by resource type. These resource types are hardware, software, human labor and indirect costs. There will be a fifth subsection regarding contingency and budget control. At the end of the section there will appear the final sum of the budget for this project.

5.1.1 Hardware

The following table contains the budget for the hardware resources:

Product	Cost	Useful Life	Amortization
Local Server	499.00	4 years	20.79
MacBook Pro	1,099.00	4 years	85.85
Total	1,498.00		106.64

Table 4: Budget for Hardware Resources

Amortization has been computed based on the number of hours of the tasks to the item's associated tasks. Taking the lifespan of the product, which is four years (roughly 7680 hours), the amount is computed as follows:

$$\frac{cost}{7680} * hours\_of\_use$$

5.1.2 Software

The following table contains the budget for the software resources:

Product	Price	Useful Life	Amortization
Mac OS X 1.9.5	19.95	3 years	2.07
Matlab & SimuLink	2,150.00	3 years	223.95
Sublime Text 2	0.00	3 years	0.00
LaTeX	0.00	3 years	0.00
Total	2,169.95		226.02

Table 5: Budget for Software Resources

5.1.3 Human resources

This subsection details the budget allocated to human resources. It has been computed on the basis of hours of work per role.

Task	Hours	Project Manager	Analyst	Programmer	Cost per task
T1	25	50%	50%	0%	$1,125.00
T2	80	60%	20%	20%	$3,440.00
T3	50	10%	10%	80%	$1,450.00
T4	80	10%	60%	30%	$2,920.00
T5	80	60%	40%	0%	$2,760.00
T6	180	10%	60%	40%	$7,020.00
T7	60	10%	50%	40%	$1,400.00
T8	25	50%	50%	0%	$2,700.00
T9	75	40%	30%	30%	$987.50
TOTAL (hours)	680	158.5	321	230.5	**$23,802.5**

Table 6: Number of hours per role

Role	**Hours**	**Cost per hour**	**Cost**
Project Manager	171.5	$50.00	$8,575.00
Analyst	271	$40.00	$10,840.00
Programmer	175.5	$25.00	$4,387,50
Total			**$23,802.50**

Table 7: Total cost of internal human resources (Standard rates)

In addition to the costs generated by the work hours of the members of the team, we need to take into account how much we will be paying to the subjects on Amazon Mechanical Turk who will be taking part in our experiment.

We will need 30 subjects per condition, and we will be analyzing at least four conditions (control + three levels of manipulation). We have decided to pay our subjects $10.00 per experiment. Assuming we need at least 100 subjects, the total cost of the experiments is **$1,000.00.**

5.1.4 Indirect costs

Here we specify the costs derived from the working environment such as electricity and Internet access. Paper has not been included since all the data and reports are going to be submitted in a digital form. Costs are calculated following standard prices for the State of Colorado.

The quantity of kWh has been estimated using the consumption per hour of our electrical devices and multiplied by the number of hours each device is used and divided by 1000.

Device	Power	Hours	kWh
MacBook	80W	600	48
Server	500W	320	160
Lights	2x40W	600	48
Total			**256**

Table 5: Energy consumption by device

Product	Price	Quantity	Total
Electricity	0.089 $/kWh	256 kWh	22.79
Internet Provider	35.99 $/month	4 months	143.96
Total			**166.74**

Table 6: Budget for indirect costs

5.1.5 Unforeseen costs

As it has been mentioned in previous sections, there is a chance we might need to run more experiment to test more conditions, depending on how conclusive the results are. For this reason we have planned a special unforeseen costs budget to pay for the subjects of the experiment and also for human resources.

The unforeseen costs are calculated depending on their probability of happening. We deem there is a 30% chance we might have to run more experiments. This means that there is a 30% chance we might have to repeat tasks T6 and T7 ($7,020.00 and $1,400.00 respectively).

Total for unforeseen costs: $2,526.00

5.1.6 Contingency

We have decided to add an extra 10% of the subtotal of the budget for contingency reasons.

Total for contingency: $2,682.79

4.1.7 Total

The following table shows a summary of the different costs and the total estimate.

Concept	Cost
Hardware	106.64
Software	226.02
Human Resources	23,802.50
Experimental Subjects	1,000.00
Indirect costs	166.74
Unforeseen costs	2,566.00
Contingency	2,286.79
Total	**$30,654.69**

Table 7: Total Budget

Despite the changes produced in the temporal planning throughout the development of the project, the total cost of the project remains the same.

6 Sustainability

As part of the degree in computer science we are required to learn and apply certain transversal competences. One of these competences or skills is Sustainability. This is the reason why there is a sustainability section in this project. Here we analyze and evaluate sustainability aspects of this project regarding three different fields: economy, society and environment.

6.1 Economic sustainability

Although this project is primarily focused to academia and profit is not among its objectives, it will be assessed as if it were going to be a commercial product.

In previous sections of this document we can find budget assessment of both human and material resources.

The total cost indicated under the budget section, seems very reasonable. This is because a major proportion of the budget is allocated to human resources. The salary of the members of the team is not higher than market price in the Greater Denver Area. Hence, we are keeping the cost of human resources in check.

The only way this project could cost less is if took a shorter time to complete. Because of the extensive initial research, we can assure that we are not wasting time developing existing algorithms or technologies that we want to include in our project. As it was mentioned in the *Time Planning* section, we are giving ourselves a margin to perform more experiments. The costs of running extra experiments are contemplated under unforeseen costs. If we end up not needing to perform these experiments we will be completing our project in the minimum possible time.

The scope of this project does not include further maintenance or updates of the software produced. For this reason, these costs have not been contemplated in the budgeting.

It is not clear whether this project would be viable if it were intended for commercial purposes. To break even, we would have to sell our tool for at least the price of the total budget or use advertising or other means of profit. However, if this were the case we would probably need to contemplate maintenance costs as well.

During the first stage of the project we have invited to discuss our project and consulted with members of other fields of research, such as professors in the department of psychology and cognitive science at the University of Colorado Boulder. However, no formal collaboration has been established.

This project has been awarded a 7 in economical sustainability.

6.2 Social sustainability

This project is being the developed in the academic environment of the University of Colorado Boulder. This university produces a large quantity of knowledge every year. The results of our experiments will be a contribution to science and will be available for other researchers to use. This is the reason why this project improves the situation of our current environment.

If our project completes all the objectives successfully, our system may be integrated in learning tools. Anyone wanting to acquire in-depth knowledge of a number of topics (planes, birds, plants, etc.) may benefit from our system, achieving faster learning rates. The purpose of our tool is to boost performance when categorizing objects. The users will benefit from, making their learning process easier and faster.

Teachers and professors may also benefit prom this tool since they will be able to use it to pace their students in the right manner.

Furthermore, if our system proves itself valuable there could even be lifestyle applications to it. We could possibly find a way to improve the visual experience of people with visual disabilities using our system.

We do not think our project will suppose a disadvantage for any collective of people. We believe this because we are not launching any commercial product and our system does not intend to substitute any existing commercial products. Also, our results will be at the disposal of any collective who wants to make use of them.

This project has been awarded a 9 in social sustainability.

6.3 Environmental sustainability

The material resources allocated to each task were specified in the previous document (*Temporal Planning*): Computer is assigned to all tasks; local server is assigned to experimental tasks.

Assuming the computer will be working during the entirety of the project (600 hours) at 80 W and the local server will be working at every time during experimental tasks (320 hours) at 500 W and two fluorescent lights working for 600 hours at 40 W each, the total of our power consumption is 256 kWh. This sum translates to emissions of 110.94 kg of CO_2. Although this is a relatively high figure, this is the maximum possible. Since neither the server nor the computer will be running at all times during their associated tasks. The only way to minimize the volume of CO_2 emissions in this project is by replacing our resources with more energy efficient equipment. However, the environmental costs of replacing equipment would have to be assessed.

If this project were not a part of my Bachelor's Thesis the environmental impact would probably be the same, being our workplace an academic environment, the procedures are the same for any kind of project.

In terms of recycling previous knowledge, we have done an extensive research of visual perception and decision-making processes. We will be using an existing dataset of a previous study. This saves us nearly a month of collecting data to predict diagnostic regions. For this reason, we are avoiding unnecessary costs, time and energy.

As stated in the previous section, the result of our work will be at the disposal of other researchers. Therefore, our project will help reduce time, costs and consumption of energy in the future.

This project has been awarded an 8 in environmental sustainability

7 Legal aspects of the project

There were three important legal aspects to be considered when developing the project: IRB approval, software licenses and plagiarism.

7.1 IRB Approval

This research project involves human subjects. By U.S. federal law, any research that requires the investigator to interact with a human being requires prior approval by the Institutional Review Board (IRB), in this case the University of Colorado's IRB.

Before undertaking any such project, at least one person in the team must complete an online tutorial in the protection of human research subjects through the CITI Program. The tutorial will take 45-60 minutes to complete, and the certificate of completion is valid for three years.

It is very important to comply with this regulation. My advisor's IRB approval is sufficient for us to be allowed to conduct experiments involving humans.

7.2 Software Licenses

All the software tools that have been used during the development of this project are either open source or the University has provided a license for their use.

7.3 Plagiarism

It might seem obvious that plagiarism is in some cases illegal and in all cases a bad practice. However, one may not be aware that certain practices are indeed plagiarism even when it is not intended.

Research involves the use of other people's work and it is important to know how to use this information and how to properly reference your sources.

At the beginning of my stay the university provided me with a handbook on good practices for writing for research. I had to read this handbook and use these techniques when writing reports and other documents.

CHAPTER III. EXECUTION

On this chapter of the thesis I will describe how the project was executed. This chapter involves the initial stages of the project in which we decided on the data set and the manipulations, the design and implementation of the final experiment and a summary of the results we obtained.

1 Collection of initial data (CUB-200 dataset)

At the early stages of the project we had to decide how exactly we planned to provide an answer to our main question. To pursue our objective, we required a comprehensive image dataset that contained several examples of many different categories.

We decided to use the dataset of bird images collected by Welinder et al. in a technical report called **Caltech-UCSD Birds 200** [10]**.** This dataset contains 6,033 images of 200 species of birds (200 categories). Each bird species has an associated vector of 287 binary visual attributes such as “has stripped head” or “has red eyes”.

Compared to basic level recognition, fine-grained categorization can be more challenging, as there are in general less data and fewer discriminative features. This necessitates the use of stronger prior for feature selection. In this work, we include humans in the loop to help computers select discriminative features.

The most important reason why we chose to utilize this data is that Deng et al. had already conducted a study using this image dataset. In this study, Deng introduces an online called “Bubbles” that reveals discriminative or diagnostic features that humans use to categorize visual stimuli.

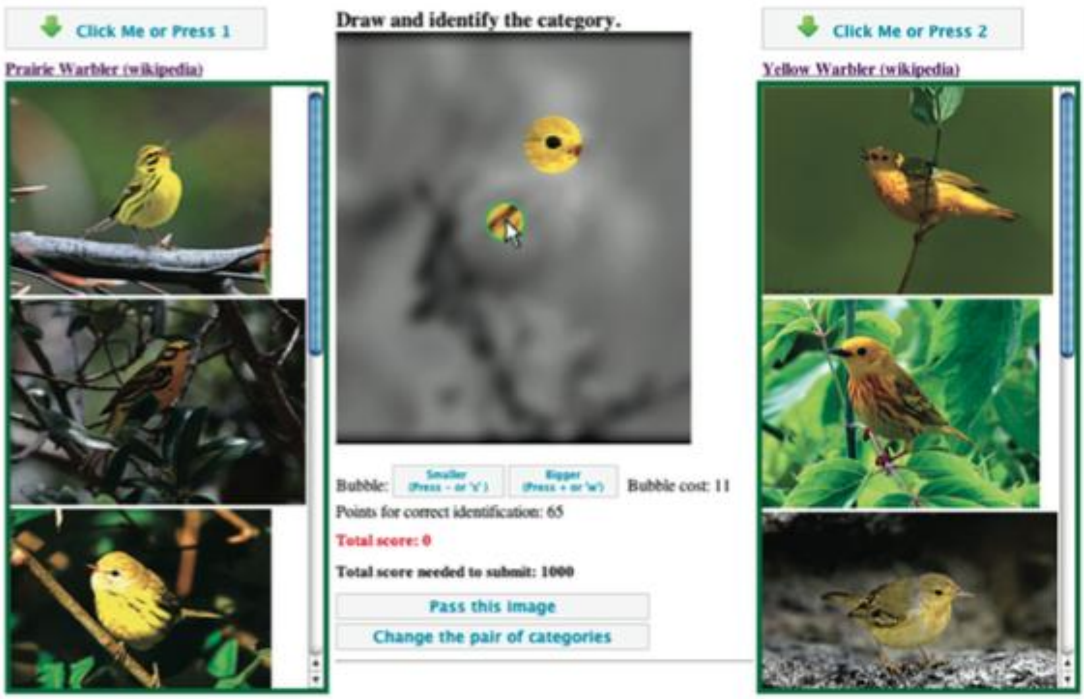


Figure 4 Screenshot of the Bubbles Game [REF]

The player’s goal is to identify the category of a heavily blurred image. During this game, the player can choose to reveal certain details of a circular region [12] The more you reveal the less points you get for your answers, assuring thus the quality of the information

Because they use the CUB-200 dataset for this game, we were able to carry out our study without having to collect data on the diagnostic features of the image dataset.

Using the information on discriminative features the collected by this game we want to create heat maps of the areas of the image that users found most useful at the time of categorizing images. We will then use these heat maps to apply image manipulations accordingly in order to highlight certain features of an image to which we want to direct people's attention.

2 Extracting diagnostic regions

In order to be able to apply any manipulation we would come up with to highlight diagnostic regions, we first had to extract these regions from the information we had.

Before I go any further explaining our diagnostic regions it is important to introduce the concept of a trial in our experiments. Trials consist of a test image, a reference image A, and a reference image B. References A and B belong to different categories. The test image belongs either the category of reference A or the category of reference B.

Figure 5 Structure of a trial

For the purpose of our task, we want to manipulate reference images as opposed to the test image. Deng's study contains information on diagnostic regions of the test image. This is the reason why we need to carefully craft our heat maps so that they provide valuable information.

For example, we want to present a trial that shows a reference image A that belongs to category 1 and a reference image B that belongs to category 2. To obtain diagnostic the diagnostic regions for reference image A, we can only use the trials in the bubbles games in which reference image A was the test image and was compared against images in category 2. To obtain the diagnostic regions of reference image B we would use the trials in which this image was the test image in the bubble game and was compared against images of category 1.

Figure 6 Example of image in the dataset

For every player who played a certain trial in the bubbles game we know the area they uncovered. We create a logical matrix of the size of the image and set to one the pixels inside the uncovered area.

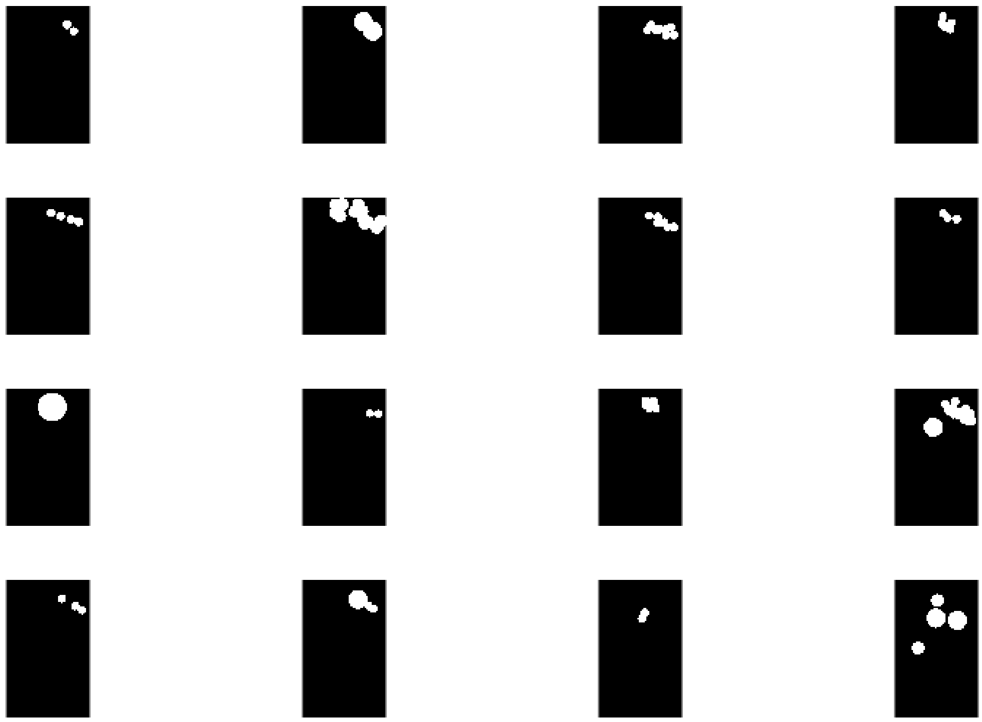


Figure 7 Uncovered areas of 16 different players

These matrices contain a logical value 0 or 1. Nevertheless, because of the way the game was designed, we know that the higher the score for a player the more information provides the uncovered area. Hence, we normalize the scores and multiply each of these matrices by their respective player's score.

$$HM = \sum Score_p * HM_p$$

After this, HM is divided by its maximum value so that every value is between 0 and 1. In order to provide a more natural feel when using the heat map as a mask for the manipulations, we apply a Gaussian filter to the matrix, making the steps smoother. This is what the final form of the heat map looks like:

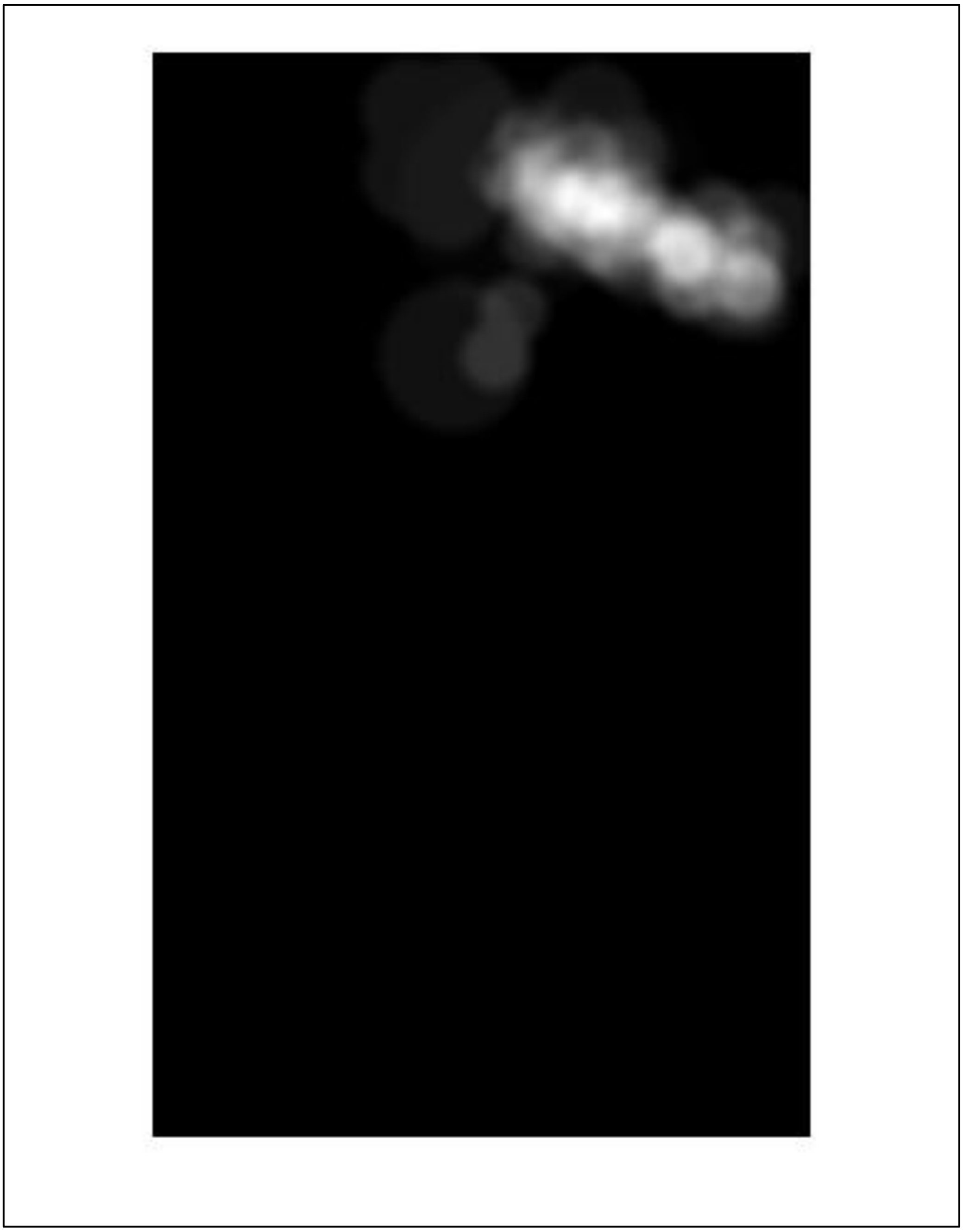


Figure 8 Final heat map of diagnostic regions

3 Image Manipulations

One of the key parts of this study is how to manipulate the images so as to direct peoples' attention to the important parts of the image. At the beginning of this project there was much discussion and debate on how to proceed on this matter.

We agreed on applying two kinds of manipulations:

- Desaturation: subtracting color from non-diagnostic regions (possibly boosting saturation on diagnostic regions, too).
- Gaussian Blur (or Spatial Frequency manipulation): applying a low-pass filter to non-diagnostic regions so as to blur

The idea was to test subject on different points of the two-dimensional space of manipulations. Figure 9 shows an example of different points in which we wanted to test the efficacy of our manipulations.

However, it felt like desaturation was taking away too much information from the images. During a lab meeting we consulted the rest of the members of the lab who were not involved in this project and they agreed that desaturation was both distracting and ineffective.

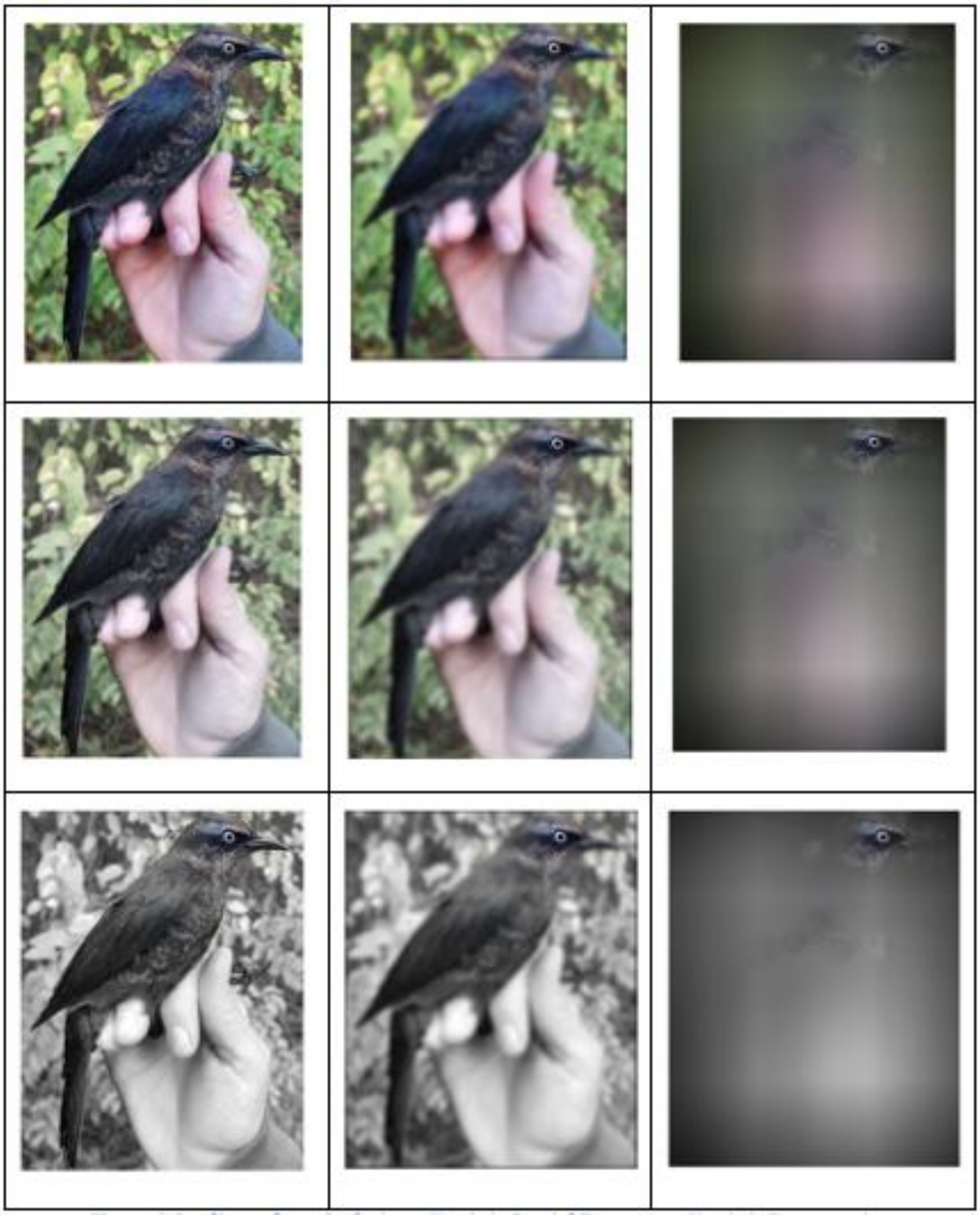


Figure 9 Gradient of manipulations: X axis is Spatial Frequency, Y axis is Desaturation

At this point we had to reassess the direction of our research in order to avoid spending time and resources on something that did not look very promising.

In this regard, we decided to go ahead and try whether attentional highlighting could work with only a spatial frequency manipulation. We generated a gradient of different levels of manipulations using the sigma parameter of the Gaussian blur equation.

$$G(x,y) = \frac{1}{2\pi\sigma^2} * e^{-\frac{x^2+y^2}{2\sigma^2}} \quad [12]$$

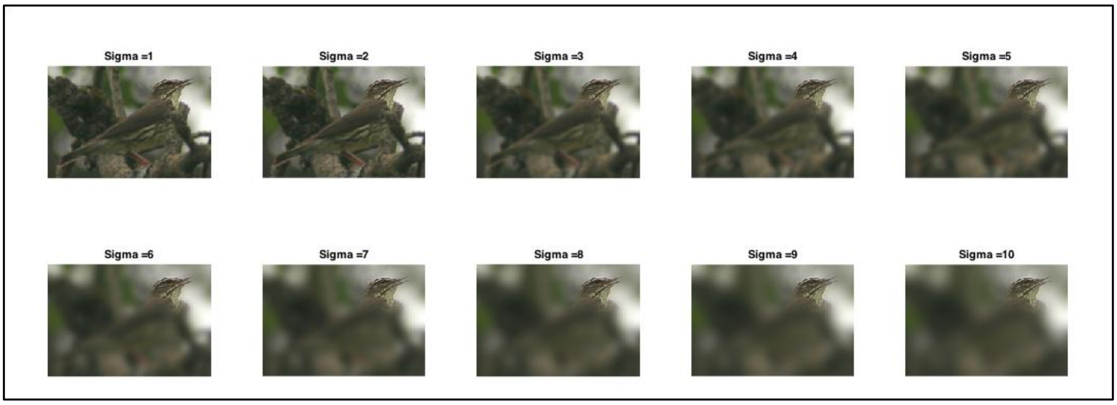


Figure 10 Gradient of different levels of spatial frequency manipulations

We deemed appropriate to choose an intermediate level of manipulation. For this reason we decided to use σ = 5.

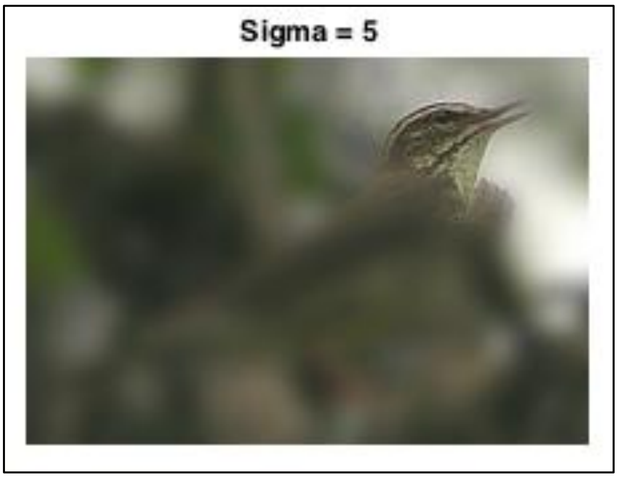


Figure 11 Example of manipulated image. Sigma = 5

4 Similarity of birds

Once we had decided to only use the spatial frequency manipulation on the images, and after having had a closer contact with the image, we came up with the idea of classifying trials according to difficulty.

The way we measure difficulty is by how similar the two categories of birds included in trial are. To define the concept of similarity between two images of birds we rely on two dimensions:

- Biological family
- Visual attributes

On the CUB-200 we have the necessary information to classify images according to these parameters.

For the **biological family** dimension, we generate a binary matrix that represents whether a pair of categories of birds belongs to the same biological family.

Category	1	2	3	...	N
1	1	1	0	...	0
2	1	1	1	....	1
3	0	1	1	...	0
...	...	...	...	....	...
N	0	1	0	...	1

Figure 12 Matrix of Biological Family

In order to compute the similarity in terms of **visual attributes,** we use a set of 287 binary visual attributes (e.g., has blue crown, has a stripped head, has red eyes). For every pair of categories of birds, we compute the cosine similarity of their visual attributes vector.

The cosine similarity is calculated as follows:

$$simil(x, y) = \cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \times \|\vec{y}\|} \text{ [12] [13]}$$

To create the four levels of difficulty we combine these two measures of similarity:

- **Within-Family / Visually Similar (W.S.)**: Top ranked pairs by visual attribute cosine similarity that belong to the same biological family.
- **Within-Family / Visually Dissimilar (W.D.)**: Lowest ranked pairs by visual attribute cosine similarity that belong to the same biological family.
- **Across-Family /Visually Similar (A.S.):** Top ranked pairs by visual attribute cosine similarity that belong to different biological families.
- **Across-Family / Visually Dissimilar (A.D.)**: Lowest ranked pairs by visual attribute cosine similarity that belong to different biological families.

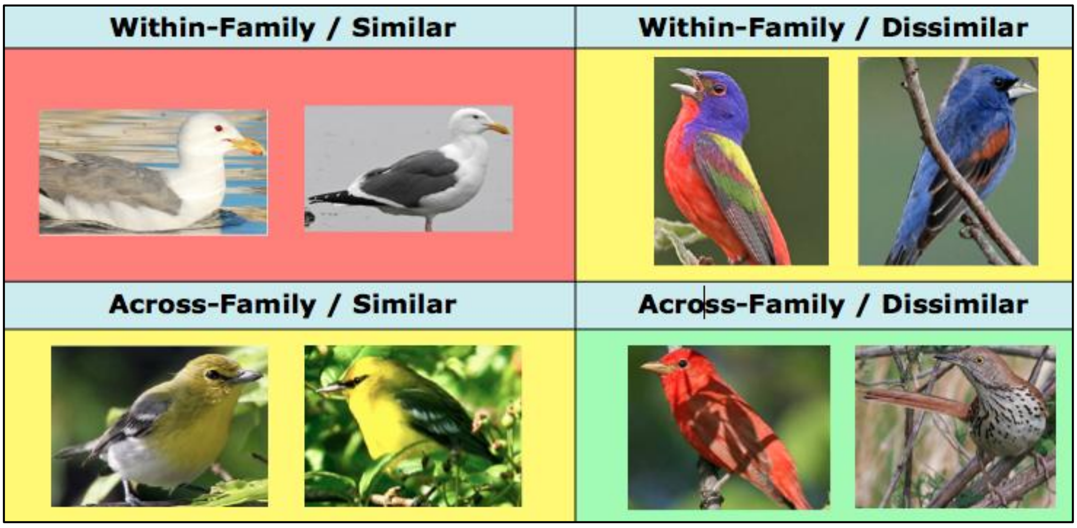


Figure 13 Diagram of trial types

5 Conditions of the final experiment

Even though the data was collected in several rounds of experiments so as to see we were going in the right direction, I will treat the rounds of experiments as one big final experiment that includes all the conditions.

There are two between-subject conditions and four across-subject conditions on this experiment

5.1 Between-subject conditions

<u>Manipulated vs. Non-manipulated reference images</u>

Fifty subjects were randomly assigned to each of the following conditions:

Figure 14 Example of a trial in the Non-manipulated condition

Figure 15 Example of trial in the Manipulated condition

Feedback vs No-Feedback

Twenty-five subjects of each of the previous conditions were assigned to an experiment in which feedback was provided. At the end of every trial they were informed of whether they had classified the bird correctly or not.

The remaining fifty subjects did not receive any kind of feedback during the entire experiment.

5.2 Across-Subject Conditions

According to the measure of similarity that was established on the previous section (Similarity of birds), we created four different types of trials.

Every subject in the experiment, regardless of his or her between-subject condition, was given forty trials in the following four conditions (according to our similarity/difficulty measure).

The following figures show an example of each of these types trials. This time our hypothesis seemed to be accurate. To the naked eye, the trials of the type "Within-Family / Visually Similar" are noticeably much more difficult than the ones of the "Across-Family / Visually dissimilar" type.

- **Within-Family / Visually Similar (W.S.)**:

Figure 16 Example of WS trial

- **Within-Family / Visually Dissimilar (W.D.)**:

Figure 17 Example of WD Trial

- **Across-Family /Visually Similar (A.S.):**

Figure 18 Example of AS Trial

- **Across-Family / Visually Dissimilar (A.D.)**:

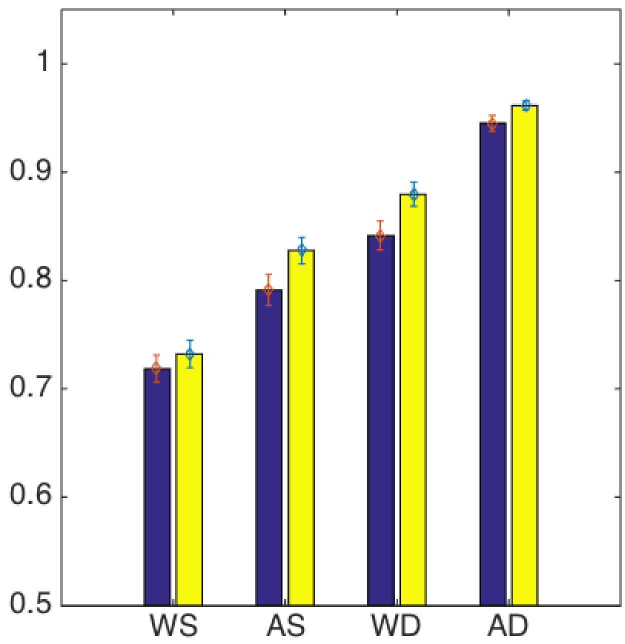
Figure 19 Example of AD Trial

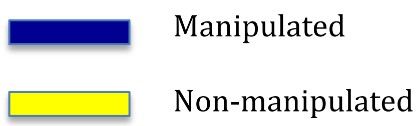
6 Results

On this section I analyze the data resulting from the final experiment. As stated at the beginning of the project, our objective is to see whether attentional highlighting has a positive effect on visual categorization. We defined this positive effect, or better performance, as an improvement in accuracy and shorter response times when compared to not using attentional highlighting.

6.1 Accuracy

The following bar chart shows the results on accuracy:





Accuracy

WS: $t(98) = 0.53$

$p = 0.60$

AS: $t(96) = 1.89$

$p = 0.06$

WD: $t(94) = 2.16$

$p = 0.03$ *

AD: $t(80) = 1.89$

$p = 0.06$

Figure 20 Mean accuracy

As can be seen in the chart, subjects on the manipulated condition performed a bit worse than those who were on the control experiment. However, this difference is not statistically significant in most of the within subject conditions.

7.2 Log response time

The following char shows the mean log response times in *ms* for every condition of the subject

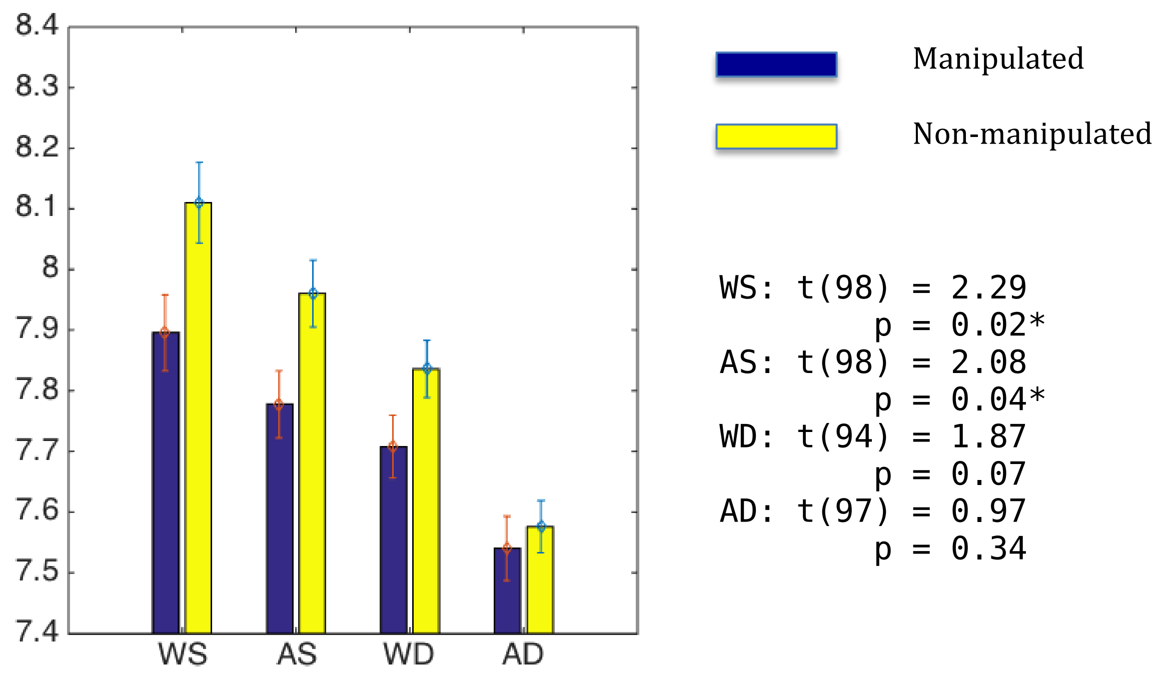


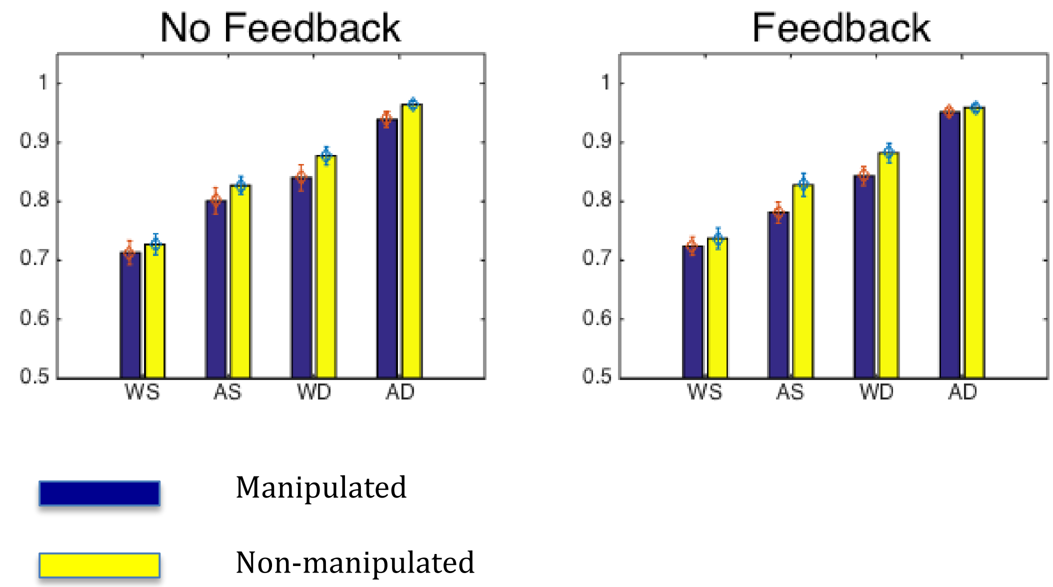
Figure 21 Mean Log Response Time

These results show that response times were shorter for the subjects on the condition that used attentional highlighting. We do not see significant results in every within-subject condition, but only on the most difficult trials, this is, those that included the pairs of categories of birds with highest visual attribute similarity (WS and AS).

7.1 Feedback vs. No Feedback

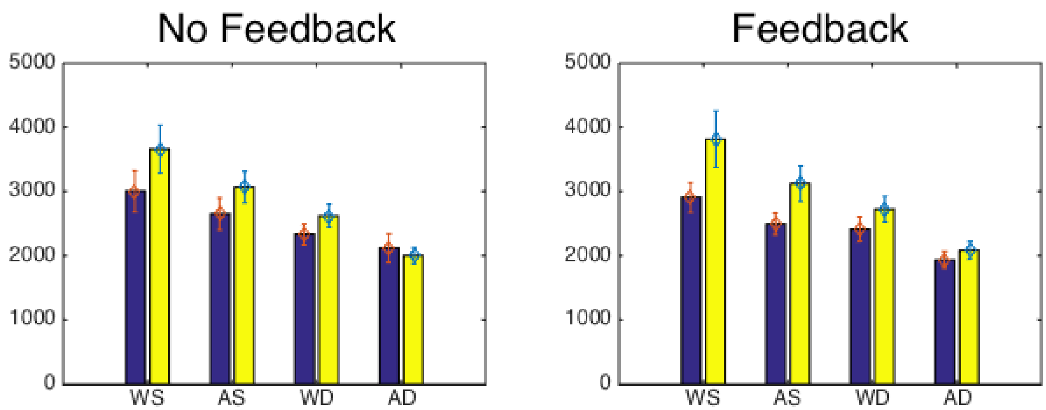
We analyzed whether feedback had an effect in overall accuracy or not. As we can extract from the following graphs, accuracy remains statistically the same regardless of whether subjects were give feedback after every trial.

ACCURACY



We did the same for response time:

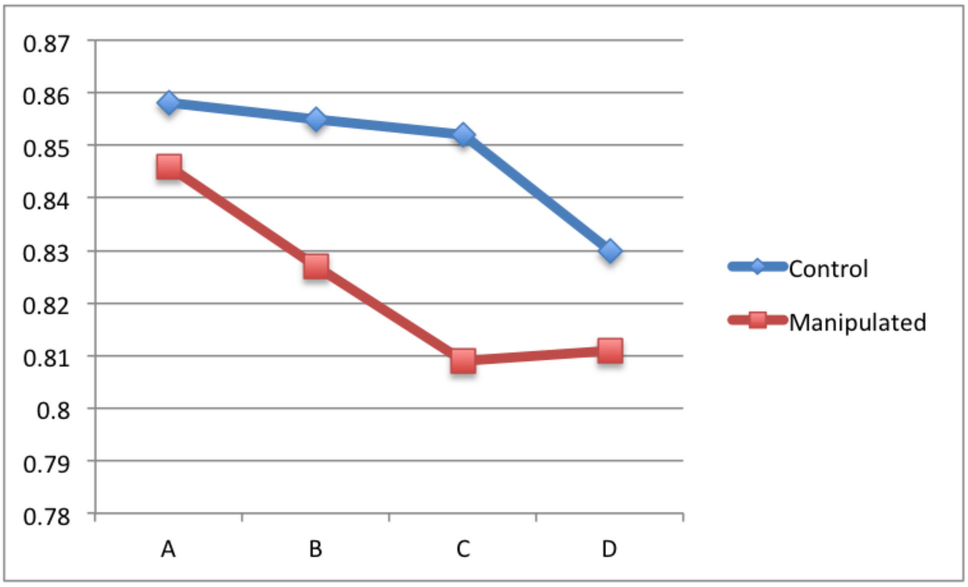
RESPONSE TIME



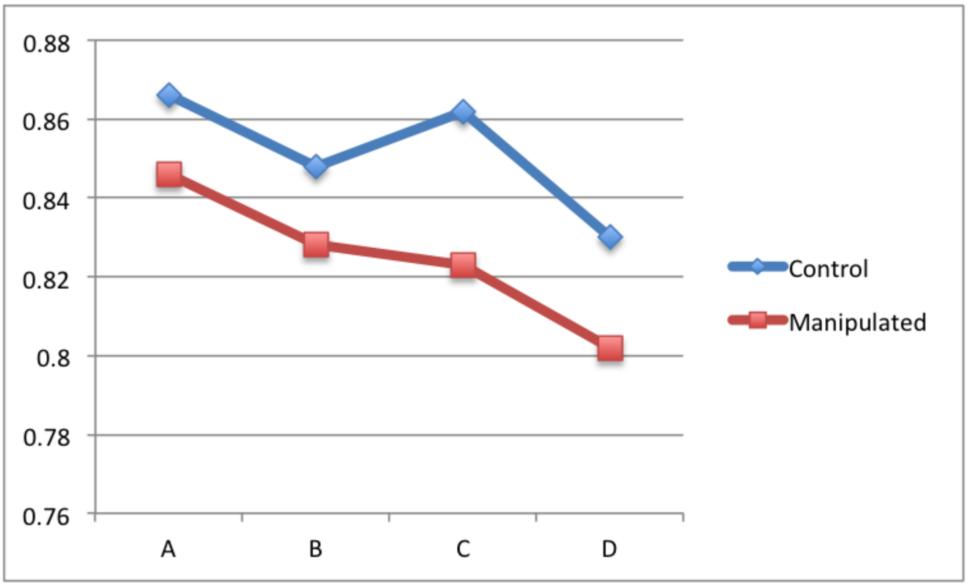
Again, the difference between the two conditions was not significant.

What we really wanted to find by using the feedback vs. no feedback condition is whether there would be a "training" effect on the subjects that were given feedback. For this reason, the experiments were divided into four blocks (A, B, C, D). Every block contained the same amount of trials of each difficulty level.

ACCURACY PER BLOCKS WITHOUT FEEDBACK



ACCURACY PER BLOCKS WITH FEEDBACK



Accuracy does in fact decrease from block to block. These results seem to suggest that, even when they were receiving feedback, they were not learning to perform better from block to block.

Although it is discouraging to find no training effect at all, there might be other explanations for these results. One reason could be that subjects may lose interest in the task because of it is monotonous. Another reason might be that we may have not allowed enough time for them to properly learn the task.

CHAPTER IV. CONCLUSION

This project was set out to answer the question of whether attentional highlighting can be used to boost performance of novices when carrying out visual categorization tasks.

In this study we explore the concept of attentional highlighting by means of research and trial and error. We examined the application of different kinds image manipulation techniques using an intelligent scheme to extract diagnostic regions.

We designed and run a series of experiments to try out these image manipulation techniques on visual categorization tasks. We had one hundred subjects take part in a study where compared accuracy and response time across the between-subject conditions and across within-subject trial types.

Results

Accuracy did not vary across the two conditions. Attentional highlighting decreased response times by approximately 750 ms (18%) for the hardest trial type. These results suggest that attentional highlighting can be used to reduce the temporal costs associated with making difficult visual category distinctions.

Implications

Counter to our initial hypothesis, attentional highlighting does not significantly boost novice discrimination ability.

We found some promising results that these types of techniques could be apply to enhance the response times in difficult visual categorization tasks. Further experimentation is necessary to find a clearer threshold of difficulty for which attentional highlighting is the right discrimination aid.

Our experiments were not able to prove that there is a training effect when we aid individuals in a visual categorization task using attentional highlighting and providing feedback. As it was mentioned in the results this might be due to the limitations of the study.

We were able to prove that there is a benefit in visual categorization time but accuracy did not improve at all. In fact, the data suggests that there might be a trade off in speed versus accuracy.

Limitations and future direction

This study, nevertheless, was subject to certain limitations and there is room for improvement in future studies.

Firstly, one of the limitations in our study was the way we extracted diagnostic regions. By relying on Deng's Bubbles game data, we conferred users of all levels of expertise the ability to produce diagnostic regions. In future studies, experts in the field could be assigned to provide information on diagnostic regions.

Secondly, image manipulations utilized to generate attentional highlighting. It is a possibility that other image manipulations (i.e. circling diagnostic features) may serve the purpose of attentional highlighting better than the ones we were using.

Finally, one suggestion for future studies could be setting a time limit for every trial. Our experiments did not have a time limit per trial, it could be the case that that extra pressure of a timer may force subjects to make quicker decisions in which attentional highlighting may play a larger role.

Sustainability

Even if part of the results of this study, did not confirm our initial hypothesis. They may serve future researchers in the field. As mentioned in the sustainability section of the management chapter, this project intends to generate and disseminate knowledge. This objective has been achieved beyond expectations when I was invited to present a poster with our findings at the iSLC Conference 2015 at the University of California San Diego.

Conclusion

In summary, our results suggest that there is a positive effect in response times when applying attentional highlighting to visual categorization tasks. This leaves the door open to more research and experimentation since there are still more techniques to be tested and different experiments to be run on this field. We are proud of our achievements and of having been able to make our contribution to knowledge and science.

REFERENCES

[1] **Wikipedia**, *"Decision-making",* Wikipedia – The Free Encyclopedia, 2015. Online Resource. [last visited on March 14 2015] Available at: http://en.wikipedia.org/wiki/Decision-making

[2] **Wikipedia**, *"Visual Perception"*, Wikipedia – The Free Encyclopedia, 2015. Online Resource. [last visited on March 14 2015] Available at: http://en.wikipedia.org/wiki/Visual_perception

[3] **Berg, T.; Belhumeur, P.N**, "*How Do You Tell a Blackbird from a Crow?"*. 2013. In: Computer Vision (ICCV) 2013, pp 9-16 .Online Resource [last visited on March 14 2015] Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6751110&tag=1

[6] **Kumar, N.; Belhumeur, P.; et al.,** *"Leafsnap: A Computer Vision System for Automatic Plant Species Identification",* 2012. In: Computer Vision – ECCV 2012, pp 502-516 .Online Resource [last visited on March 14 2015] Available at: http://link.springer.com/chapter/10.1007/978-3-642-33709-3_36

[5] **Rasmussen, C.; Williams, C.K.I.** *"Gaussian Processes for Machine Learning",* The MIT Press, 2006.

[6] **Itti, L.; Koch, C.** *"Computation Modelling of Visual Attention",* MacMillan Magazines Ltd, 2001. In: Nature Reviews, Vol. 2. No. 3, pp194-203

[9] **Zhang, L; et al.** *"SUN: A Bayesian framework for saliency using natural statistics"* ,2008. In: Journal of Vision, Vol. 8 No. 7 Article 32 pp 1-20

[8] **Hagen, et al.**, *"The role of color in expert object recognition",* 2014. In: Journal of Vision, Vol. 14 No. 9 Article 9. Online Resource [last visited on: March 15 2015] Available at: http://www.journalofvision.org/content/14/9/9.short

[9] **Richler, J. J; Palmeri, T. J.,** "*Visual category learning*". 2014. In: WIREs Cogn Sci, Vol. 5, No 1, pp 75–94. Online Resource [last visited on: May 30 2015] Available at: http://onlinelibrary.wiley.com/doi/10.1002/wcs.1268/abstract;jsessionid=6FA7B54F2F01F7B98508CC7289CBD2A3.f04t01?deniedAccessCustomisedMessage=&userIsAuthenticated=false

[10] Welinder P., Branson S., Mita T., Wah C., Schroff F., Belongie S., Perona, P. "*Caltech-UCSD Birds 200*". California Institute of Technology. CNS-TR-2010-001. 2010.

[11] **Deng, J; et al.,** *"Fine-Grained Crowdsourcing for Fine-Grained Recognition",* 2013. In: Computer Vision and Pattern Recognition (CVPR) IEEE Conference on, pp 580-587. Online Resource [last visited on May 30 2015] Available at. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6618925&tag=1&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6618925%26tag%3D1

[12] **Shapiro, L. G.; Stockman, G. C**., "*Computer Vision*", Prentice Hall 2001 pp. 137-150. Prentice Hall, 2001

[13] **Nixon, M.S.; Aguado, A.S.**, "*Feature Extraction and Image Processing*"., Academic Press, 2008, p. 88