

Designing a Unity Plugin to Predict Expected Affect in Games Using Biophilia

by

Licheng Zhang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2022

© Licheng Zhang 2022

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Video games can generate different emotional states and affective reactions, but it can sometimes be difficult for a game's visual designer to predict the emotional response a player might experience when designing a game or game scene. In this thesis, I conducted a study to collect emotional responses to video game images. I then used that data to both confirm past research that suggests images can be used to predict affect and to build a model for predicting emotion that is specific to games. I built both a linear regression model and three neural network models to predict affective response and found that the neural net that leveraged ResNet-50 was most effective. I then incorporated that model into a Unity plug-in so that designers can use it to predict affect of players in real time.

Acknowledgements

I would like to thank my supervisor, Mark Hancock, for his continuous support. I would also like to thank Deltcho Valtchanov, Jim Wallace and Adrian Reetz who made this thesis possible.

Table of Contents

List of Figures	viii
List of Tables	xi
List of Abbreviations	xii
1 Introduction	1
1.1 Background	1
1.2 Hypotheses and Research Questions	2
1.3 Method	3
1.4 Contributions	4
1.5 Thesis overview	4
2 Related Work	6
2.1 Affective Computing	6
2.2 Measuring Emotion	7
2.3 Psychology of Colour Preference and Spatial Frequency Preference	8
2.4 Convolutional Neural Networks	9
2.5 Affective Status Prediction	10
2.6 Summary	13

3	Study: Exploring the Emotional Response to Video Game Images	14
3.1	Introduction	14
3.2	Participants	15
3.3	Design	16
3.4	Materials	18
3.5	Procedure	18
3.6	Results	19
3.7	Summary	24
4	Prediction Model Using Multiple Linear Regression	25
4.1	Formulation of Multiple Linear Regression Model	25
4.2	Examination of Linearity Assumption	27
4.3	Multiple Linear Regression Results	32
4.4	Examination of Independent Variable’s Relative Importance	32
4.5	Comparison of Multiple Linear Regression Model and Valtchanov’s algorithm	34
4.6	Summary	34
5	Neural Network Models to Predict Affect	35
5.1	Training & Dataset	35
5.2	Multi-layer perception network	36
5.3	Convolutional neural network with residual neural network (ResNet)-50 . .	39
5.4	Integration of user-specific profiles	40
5.5	Performance comparison	41
5.6	Summary	44
6	Unity Plug-in to Display Real-Time Predicted Affective Status	45
6.1	How to Use	45
6.2	Engineering Pipeline	49

6.2.1	Colour Conversion	50
6.2.2	Average Weighted Histogram	50
6.2.3	Spatial Frequency's Radial Power	50
6.2.4	From Colour and Spatial Frequency to Affective Level	51
6.2.5	Customized Neural Network	51
6.3	Summary	51
7	Conclusion & Limitations	52
7.1	Thesis Contributions	52
7.2	Limitations and Future Work	53
	References	54
	APPENDICES	58
A	R Studio Outputs	59
A.1	Summery and ANOVA of 12 Images' Arousal/Pleasure Scores	59
A.2	Summery of linear regression models	62
A.3	Neural network training process	63
A.3.1	MLP	63
A.3.2	NLP + Resnet50	67
A.3.3	NLP + Resnet50 + profile	71

List of Figures

2.1	The Affective Slider	8
2.2	Mapping from image raw data to affective levels. From (Valtchanov & Hancock, 2015). Used with permission.	12
3.1	Twelve images which were seen by every participant. The expected affective valence (EAV) values shown in brackets were calculated using Equation 2.1 (Valtchanov & Hancock, 2015)	17
3.2	Sample images in the image set.	18
3.3	Sample survey question.	19
3.4	Box plots of arousal scores and pleasure scores	20
3.5	95% confidence intervals for arousal scores for the 12 images.	21
3.6	95% confidence intervals for pleasure scores for the 12 images.	21
3.7	Comparison of arousal scores, from left to right: expected values calculated using Equation 2.1, mean of all participants, mean of participants that passed the attention check, mean of participants that failed the attention check.	23
3.8	Comparison of pleasure scores, from left to right: expected values calculated using Equation 2.1, mean of all participants, mean of participants that passed the attention check, mean of participants that failed the attention check.	23
4.1	Residuals vs Fitted plot of the arousal/pleasure model. The locally weighted scatterplot smoothing (LOWESS) (red line) in the graphs of residuals vs. fitted data is relatively flat, which provides evidence for the linearity assumption.	28

4.2	Normal Q-Q plot of the arousal/pleasure model. The points form a roughly straight line that follows the diagonal. This means the residual errors are roughly normally distributed and provides evidence for the linearity assumption.	29
4.3	Scale-Location plot of the arousal/pleasure model. The LOWESS (red line) in the graphs of Scale-Location is relatively flat, which provides evidence for the linearity assumption.	30
4.4	Residuals vs Leverage plot of the arousal/pleasure model. The spread of standardized residuals int both graphs do not change as a function of leverage (the LOWESS (red line) is relatively flat), which provides evidence for the linearity assumption.	31
5.1	Structure of the multi-layer perception (MLP) network for predicting arousal/pleasure score: (a) shows the general structure and (b) shows the details of hidden layers, including numbers of inputs/outputs and activation functions.	38
5.2	Structure of the MLP network combined with ResNet-50.	39
5.3	Structure of MLP network combined with ResNet-50 and user-profile integration.	41
6.1	The affective slider of the Unity plugin	45
6.2	Configuration window of the Unity plugin	47
6.3	Screenshot of using the Unity plug-in (a)	47
6.4	Screenshot of using the Unity plug-in (b)	48
6.5	Engineering pipeline of the Unity plugin (details are explained in the indicated subsections).	49
A.1	Training process of MLP network for predicting pleasure score (smoothing value = 0.5).	63
A.2	Validation process of MLP network for predicting pleasure score (smoothing value = 0.5).	64
A.3	Test process of MLP network for predicting pleasure score (smoothing value = 0.5).	64
A.4	Training process of MLP network for predicting arousal score (smoothing value = 0.5).	65

A.5	Validation process of MLP network for predicting arousal score (smoothing value = 0.5).	65
A.6	Test process of MLP network for predicting arousal score (smoothing value = 0.5).	66
A.7	Training process of MLP network combined with Resnet50 for predicting pleasure score (smoothing value = 0.5).	67
A.8	Validation process of MLP network combined with Resnet50 for predicting pleasure score (smoothing value = 0.5).	68
A.9	Test process of MLP network combined with Resnet50 for predicting pleasure score (smoothing value = 0.5).	68
A.10	Training process of MLP network combined with Resnet50 for predicting arousal score (smoothing value = 0.5).	69
A.11	Validation process of MLP network combined with Resnet50 for predicting arousal score (smoothing value = 0.5).	69
A.12	Test process of MLP network combined with Resnet50 for predicting arousal score (smoothing value = 0.5).	70
A.13	Training process of MLP network combined with Resnet50 and user-profile integration for predicting pleasure score (smoothing value = 0.5).	71
A.14	Validation process of MLP network combined with Resnet50 and user-profile integration for predicting pleasure score (smoothing value = 0.5).	72
A.15	Test process of MLP network combined with Resnet50 and user-profile integration for predicting pleasure score (smoothing value = 0.5).	72
A.16	Training process of MLP network combined with Resnet50 and user-profile integration for predicting arousal score (smoothing value = 0.5).	73
A.17	Validation process of MLP network combined with Resnet50 and user-profile integration for predicting arousal score (smoothing value = 0.5).	73
A.18	Test process of MLP network combined with Resnet50 and user-profile integration for predicting arousal score (smoothing value = 0.5).	74

List of Tables

3.1	Number of images where the expected score is within the confidence interval.	22
4.1	Independent variable's relative importance in arousal/pleasure model . . .	33
4.2	Comparison of Multiple Linear Regression Model and Valtchanov (2013)'s algorithm	34
5.1	Comparison of different neural network models to predict pleasure score . .	41
5.2	Comparison of different neural network models to predict arousal score . .	42
5.3	Comparison of Valtchanov and Hancock (2015)'s algorithm, the multiple linear regression model, and the neural network models to predict pleasure and arousal scores	42
A.1	Summery and ANOVA of 12 Images' Arousal Scores	60
A.2	Summery and ANOVA of 12 Images' Pleasure Scores	61

List of Abbreviations

ANN artificial neural network 10, 36

ANOVA Analysis of variance 20, 24

AS Affective Slider 7, 8, 45

CNN convolution neural network 6, 9, 10

DFT discrete Fourier transform 9

EAV expected affective valence viii, 10, 17, 45, 46, 52

GPS global positioning system 10

LOWESS locally weighted scatterplot smoothing viii, ix, 27, 28, 30, 31

MLP multi-layer perception ix, 36–41, 43, 51

PAD pleasure, arousal, and dominance 1

PCC Pearson correlation coefficient 34

ReLU rectified linear unit 36

ResNet residual neural network vi, ix, 10, 37, 39–43

SAM Self-Assessment Manikin 7

SD standard deviation 34

Chapter 1

Introduction

Game designers collect user feedback and determine the quality of the player experience through many methods, including questionnaires, in-game feedback, and app-store reviews (Davis et al., 2005). Most of these methods require collection of data after the game (or a prototype) is created and require players to provide feedback after playing. Game designers lack a tool that can provide immediate feedback to help predict a player’s emotional response to game designs as the game visuals are being created.

To help game designers better understand the short-term affective states of players, I present a Unity game engine plug-in that can predict the expected pleasure and arousal from players when viewing game visuals in real-time, two dimensions from the pleasure, arousal, and dominance (PAD) dimensions proposed by Russell and Mehrabian (1977).

1.1 Background

Video games are digital media embedded with extensive player engagement, which offer a broad range of interactive experiences and allow players to enjoy virtual worlds full of challenges (Calleja et al., 2016). Over the last few years, research has shown that video games play a key role in affecting people’s emotions. Some research studies suggest that video games can trigger stress (Porter & Goolkasian, 2019), while others have shown that video games reduce or manage stress (Carlier et al., 2020). Research has also shown that players react to video games with a large range of emotions, including anger, sadness, amusement and enthusiasm (Behnke et al., 2021).

Digital experiences in video games are facilitated by emotional responses. Emotions are one of the major factors that game designers need to consider when developing a game. Research shows that graphical and audio aesthetics in a game can largely influence the experience of playing (Lucio de Mattos, 2020). There are many elements in video games that can induce different emotional reactions. For example, research suggests it is an essential feature in game design to consider sound along with graphics to create “unique, immersive and rewarding gaming experiences” (Liljedahl, 2010).

Many theories have been proposed to provide guidelines to game designers to elicit specific emotions in players. For example, a psychological constructionist theory, called Conceptual Act Theory, suggests that “emotions are constructed when conceptual knowledge is applied to ever changing affective experiences” (Barrett, 2014). In video games, emotional responses are carefully designed stimulation, generated by specific playing experience, graphic content, or audio source. It is believed by some researchers that “understanding the role of emotion in creating truly immersive and believable environments is critical for game designers” (de Byl, 2015). Game designers elicit different kinds of emotions in players, through manipulating all game elements, including lighting, story telling, and the game environment (de Byl, 2015).

1.2 Hypotheses and Research Questions

Studies conducted by Russell and Mehrabian in 1977 provided evidence that three independent and bipolar dimensions, pleasure-displeasure, degree of arousal, and dominance-submissiveness, are “both necessary and sufficient to adequately define emotional states” (Russell & Mehrabian, 1977). In Russell and Mehrabian’s study, dominance accounted for only a trivial proportion of variance in the measured affective value and therefore dominance is generally given less attention compared to pleasure and arousal (Poels et al., 2012). As a result, in this thesis, pleasure and arousal are used as two dimensions for measuring player’s affective status.

Previous research has shown that the expected affective valence (EAV) of natural environments can be determined through visual information. Six factors are believed to be responsible for automatic emotional responses to scenes and environments: shape, contour, complexity, colour hue, colour saturation, and colour brightness (Valtchanov & Hancock, 2015).

I speculate that the same algorithm applies to game scenes as well. Therefore, I formulate the following hypotheses:

H1: Pleasure experienced while playing a digital game can be predicted by the game scene’s shape, contour, complexity, colour hue, colour saturation, and colour brightness.

H2: Arousal experienced while playing a digital game can be predicted by the game scene’s shape, contour, complexity, colour hue, colour saturation, and colour brightness.

There are several issues that, although not investigated in previous studies, are crucial to getting a more complete understanding of a player’s affective level and help game designers predict the expected affect in the process of game development. Therefore, I formulate the following research questions:

RQ1 How do colour and spatial frequencies in the video game’s scene affect player’s arousal level and pleasure level?

RQ2 Is it possible to use a linear regression model to predict video game players’ arousal level and pleasure level?

RQ3 Is it possible to use a neural network model to predict video game players’ arousal level and pleasure level?

RQ4 Can a tool be built for game designers that can predict a player’s arousal level and pleasure level?

1.3 Method

Thus, to investigate how emotional feedback can be affected by game graphics, I conducted an online study to collect people’s emotional feedback to game scenes. I used different methods to analyze the data and built a prediction model from it.

I developed a Unity Game Engine plug-in that can predict players’ levels of pleasure and arousal. Our prototype focuses on the emotional dimensions of pleasure and arousal. The predicted emotions are presented in the form of an Affective Slider (Betella & Verschure, 2016), which includes a pleasure slider and an arousal slider. The algorithm used in this Unity plug-in can be easily switched by game developers. I also provided our own machine learning model, trained by data collected through user studies. In this thesis, I present the

study design, data analysis, machine learning process, and the system design of the Unity plug-in.

This plug-in can analyze the current game frame and extract information on hue, luminance, saturation, and spatial frequencies, which are used to predict and present expected emotional response. It also uses the power of machine learning to determine the expected pleasure and arousal scores. Users can also choose to use their own model.

1.4 Contributions

We, therefore, provide four main contributions:

1. I present empirical results from a survey study that collected individuals' emotional reactions to game screenshots and explore possible variables that affect the satisfaction and arousal scores reported by participants. The data collected this study is valuable for other researchers who are interested in people's emotional responses to images specific to video games.
2. I use linear regression methods to construct a prediction model from the collected data. I validate the linearity assumption and measure the model's accuracy.
3. I use supervised machine learning methods to construct a prediction model from the collected data. This machine learning can be integrated into our prototype and make the predictions offer more reliable accuracy.
4. I design and develop a Unity game engine plug-in, a prototype tool to help game developers check the real-time automated analysis of the expected affective valence of the game scene. This plug-in can act as a tool to assist designers, developers, researchers, and academics to embed affective constructs within new games and understand those employed in existing ones. I provide a formal approach to understanding the role emotions play in games at all levels and bridge the gap between game design, development and technology.

1.5 Thesis overview

In this thesis, Chapter 2 situates the thesis within the current state of research in the fields of affective computing and affective status prediction. Chapter 3 covers a study I conducted in order to collect people's emotional feedback after seeing video game screenshots.

Chapter 4 and Chapter 5 analyze the collected data and train a machine learning model to predict players' emotional feedback. Chapter 6 describes a prototype tool for game designers, which is a plug-in for the Unity game engine that can display predicted quantitative data of players' emotions. Chapter 7 concludes the thesis and discusses potential future work.

Chapter 2

Related Work

This chapter provides some background for this thesis and the current state of research in related fields. It surveys previous work in affective computing, tools to assess affective status, and research on exploring human preference for colour and spatial frequency. This chapter also introduces convolution neural networks (CNNs) and previous work in affective status prediction.

2.1 Affective Computing

The field of affective computing has investigated the need to consider emotional response in interfaces for decades, both for measuring human response to computers, as well as eliciting intentional responses from users. Affective computing was first introduced by Rosalind Picard in 1998. It is an interdisciplinary field based in psychophysiology, computer science, biomedical engineering and artificial intelligence (Rosalind Picard, 1998). It has emerged as an important study to detect or even predict human emotions. There is a wide spectrum of promising applications in the field of affective computing such as virtual reality, human emotion recognition and many more. Some work has focused on applying psychological models of pleasure and arousal to categorize media along the two-dimensional space of valence (positive or negative) versus arousal (high or low) (Hanjalic, 2006). Our work could be considered to be in the field of affective computing, as it builds on past research to predict a video game player's emotional responses.

2.2 Measuring Emotion

In our research, we need a way to assess human emotion in an online study (introduced in Chapter 3). We also need a way to visually indicate the predicted affective state that a game will induce in our Unity plug-in (introduced in Chapter 6).

The circumplex model of affect suggests that all affective states can be described using two independent neurophysiological systems, one of which is valence (pleasure) and the other of which is arousal or activation. Posner et al. (2005) proposed that each emotional state can be represented as a linear combination of these two dimensions.

To measure these two emotional dimensions, many assessment tools have been proposed and widely used in studies to collect subjective affective ratings of participants (Betella & Verschure, 2016; Bradley & Lang, 1994; Watson et al., 1988). The Self-Assessment Manikin (SAM) is a pictorial scale for the measurement of pleasure, arousal and dominance (Bradley & Lang, 1994). Betella and Verschure (2016) more recently introduced the Affective Slider (AS) (shown in Figure 2.1), a digital self-reporting tool composed of two slider controls for the quick assessment of pleasure and arousal, which builds on the SAM and provides a more modern graphical way to assess affect. The AS was shown in their study to be more effective than SAM. These two sliders measure the level of pleasure and arousal from the circumplex model of affect (Posner et al., 2005). The AS is intended to not require written instructions and is displayed in black and white on purpose, in order to avoid any bias caused by colour preference (Betella & Verschure, 2016). We therefore chose the AS as the tool to both measure affect in our study, and to display the predicted affect in our Unity plug-in.

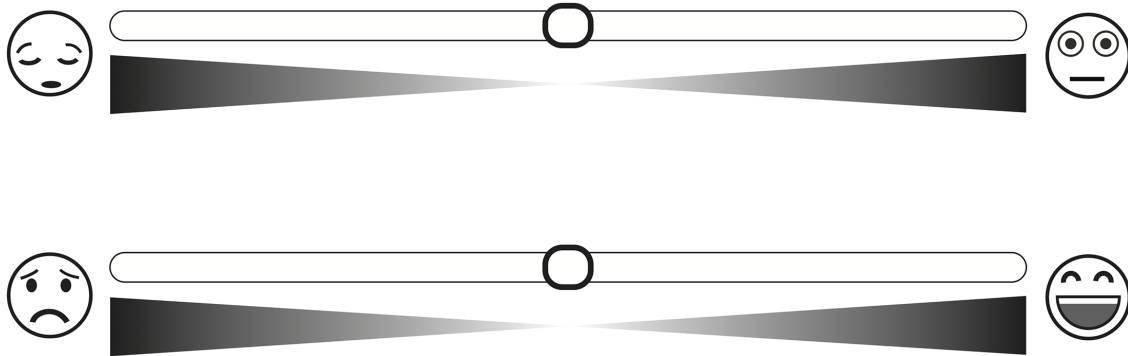


Figure 2.1: The Affective Slider

The Affective Slider (AS) is a digital self-reporting tool composed of two sliders that measure arousal (top) and pleasure (bottom) on a continuous scale. Implementations of the AS (including these images) are available here:

<http://github.com/albertobeta/AffectiveSlider>

2.3 Psychology of Colour Preference and Spatial Frequency Preference

In my research, I aim to predict human emotional responses when playing video games. In this section, I review literature on emotional response to image content. This section discusses previous research on human preference for colour and spatial frequency of the image content.

Video games have many colourful elements that may be carefully designed to elicit specific emotions for players. Colour preference is one of the significant aspects of the human visual system that affects a wide range of human behaviour, such as the choice of buying cars and the decisions made in video games. Research by Palmer and Schloss (2010) has shown that, when looking at medium-, and long-wavelength colours, both male and female participants preferred colours that were more violet to colours that were more yellow-green. However, they also found that men and women have different preferences when looking at low- and medium-wavelength colours. Women prefer redder colours while men prefer colours that are more blue-green (Palmer & Schloss, 2010). This finding suggests that personal profile may have an impact on colour preference, which were explored in this thesis.

Research has also shown that there is a significant effect of saturation and brightness on the ratings of pleasure and arousal (Wilms & Oberfeld, 2018). Research has also shown

that brighter and more saturated colour has an effect of inducing calmness. On the other hand, darker colours are more likely to cause negative emotions such as anger, hostility, and aggression (Valdez & Mehrabian, 1994).

Aside from colours, the image’s spatial frequency is another crucial factor that affects players’ emotions. The objects players see in the game scenes have different shapes, contours and complexity, which all contain different spatial frequencies that affect players’ affective status. Previous research has shown that, in terms of shape, human adults prefer curved objects to objects with sharp or pointy corners (Bar & Neta, 2006). For curvature, a similar finding has shown that non-curved/less-curved versions of objects with sharp angles and straight edges are less preferred (Carbon, 2010).

For image complexity, research has found that as the level of image complexity (density of shapes and contours) increases, the image more easily triggers negative emotions (Valtchanov, 2013).

Research suggests that the human visual system recognizes objects through spatial frequencies and spatial frequency distributions (Fintzi & Mahon, 2014). Research has also shown that different spatial frequencies have different effects on emotions. For example, Wilkins et al. (1984) found that spatial frequencies within the range of 2–8 cycles/degree are related to visual discomfort.

Fourier transforms of visual scenes allow researchers to better study the components of visual spatial frequencies and how spatial frequencies affect human emotion (Geisler, 2008). The power spectrum of an image can be calculated by taking the squared magnitude of its discrete Fourier transform (DFT) (Torralba & Oliva, 2003). Previous research has also explored the relationship between the power of spatial frequency and visual discomfort. It is suggested that images with high power at spatial frequencies to which visual system is most sensitive can trigger stress and discomfort (Torralba & Oliva, 2003).

2.4 Convolutional Neural Networks

In this thesis, convolution neural networks (CNNs) were used as a powerful tool to build an affect prediction model that was trained using data collected through an online study. It greatly boosted the model’s accuracy by incorporating more visual details of the image. More details of the implementation of CNNs are introduced in Chapter 5. In this section, I briefly introduce CNNs and how they are currently used for image analysis.

In recent years, a large amount of effort has been made to apply deep learning neural network in a wide range of areas, including computer vision and object recognition. Deep

CNNs have been widely implemented in the field of video processing, object classification and picture segmentation, natural language processing and many other areas.

Deep learning, especially CNNs, has achieved impressive performance for various computer vision and pattern recognition tasks. Like any other regular neural network, CNNs are very popular in the community of machine learning. However, they have an impressive feature that weights and biases are updated in the process of training. The goal of a CNN’s feed-forward kernel is to extract meaningful features from the training data. The output of the kernel is fed into the activation function, which only transmits the signal if its value is above certain threshold. CNNs also use loss functions and optimizers to better train the model and achieve convergence (Li et al., 2021).

A residual neural network (ResNet) is an artificial neural network (ANN) of a kind that stacks residual blocks on top of each other to form a network (He et al., 2016). For example, the popular ResNet-50 model is a CNN that is 50 layers deep. As a CNN grows deeper, the issue of vanishing gradients occurs, which negatively impacts network performance. ResNet solves this problem by including a “skip connection” feature which enables training of multiple deep layers without vanishing gradient issues (He et al., 2016). Among all the variants of ResNet, ResNet-50 has high accuracy and precision, considering its relatively small size, which makes it faster to train and easier to deploy (Manjula et al., 2022).

2.5 Affective Status Prediction

This thesis is inspired by previous research that predicts people’s affective state based on visual input. EnviroPulse is a mobile system that can automatically calculate expected affective valence (EAV) in real-time (Valtchanov & Hancock, 2015). It mainly focus on predicting human emotional response to natural environments by analyzing visual characteristics of scenes. EnviroPulse can be deployed on mobile phones or be made to work with more complex systems like global positioning systems (GPSes).

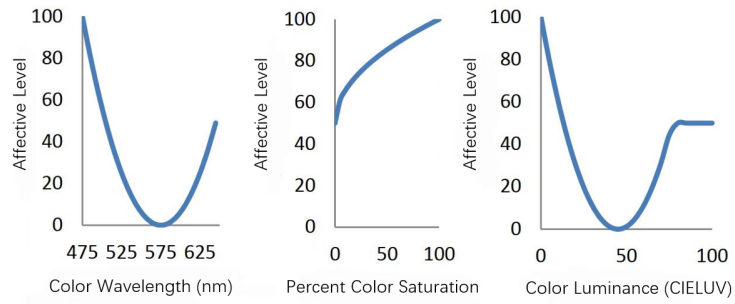
Based on Valtchanov and Hancock’s work, the EAV score can be determined by two factors, the colour score which is calculated by multiple characteristics of scenes in the colour space, and the content score which is calculated by the power of spatial frequency (introduced in section 2.3).

$$EAV \text{ Score} = 0.3 \cdot \text{colour score} + 0.7 \cdot \text{content score} \quad (2.1)$$

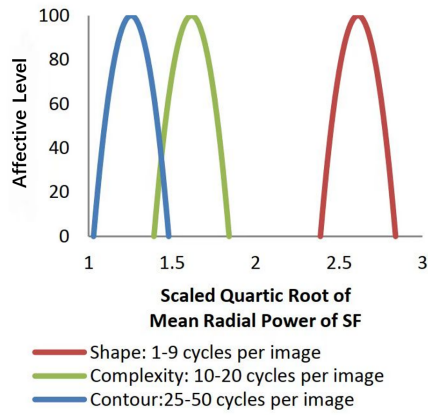
The colour score can be calculated by colour hue, colour saturation, and luminance. To make these three factors fit in a linear regression model, Valtchanov used algorithms

introduced in previous research (Fernandez & Wilkins, 2008; Juricevic et al., 2010; Torralba & Oliva, 2001; Valtchanov, 2013) to map them from colour space to affective levels. The non-linear relations between colour hue, colour saturation, colour luminance and their affective levels are shown in Figure 2.2a. In principle, the pleasure score and arousal score can be expressed as a linear combination of image hue, luminance, saturation, and an image's spatial frequencies.

People have different preferences for shape, complexity, and contour with the spatial frequencies of an image. Similarly, Valtchanov and Hancock used findings from previous research to map the radial power of spatial frequencies to affective levels as well (Fernandez & Wilkins, 2008; Juricevic et al., 2010; Torralba & Oliva, 2001; Valtchanov, 2013). The non-linear relations between shape, complexity, contour and their affective levels are shown in Figure 2.2b.



(a) Curves of mappings used in our algorithm from colour space to affective levels.



(b) Curves informed by psychology research relating affective responses to shape, complexity and contour with the spatial frequencies of an image.

Figure 2.2: Mapping from image raw data to affective levels. From (Valtchanov & Hancock, 2015). Used with permission.

2.6 Summary

This chapter discusses background knowledge that was used in this thesis to build a Unity plug-in to predict video game players' affective state. In this thesis, we built on the work done by Valtchanov and Hancock (2015) by extending the model to use a neural network and compare performance between their algorithm, a newly calculated linear regression model, and this neural network. We then incorporated the neural network model into a Unity plugin so that game designers can evaluate affective response to visual graphics as they design and build games.

Chapter 3

Study: Exploring the Emotional Response to Video Game Images

In this chapter, I describe a study I used to gather data about people's expected emotional response to video game images. Section 3.1 describes the research topic and hypotheses and the support for these hypotheses. Section 3.2 introduces participant recruitment and basic information of participants. Section 3.3 describes the study design, including what independent variables were manipulated and what dependent variables were measured. Section 3.4 describes the image data set that was used in the study. Section 3.5 describes the study procedure and what tasks participants were required to complete. Finally, section 3.6 describes the study results and analyzes the collected data. Twelve images of video games were shown to all participants and are analyzed in this chapter to compare with expected affective levels. Another 1800 images were presented to participants, but only rated by one of them. These images are discussed in the study design in this chapter, but analyzed using linear regression in Chapter 4 and used to build a machine learning prediction model in Chapter 5.

3.1 Introduction

The purpose of this research is to analyze how a video game's scene can affect a player's arousal and pleasure and to build a game engine plug-in to predict a player's affective level. I used linear regression and machine learning as the main tools for predictive modelling. Since a large amount of training data is required for linear regression and machine learning,

the purpose of this study is to gather data about people’s expected emotional response to video game images. During the COVID-19 pandemic, when in-person study was hard to achieve, an online study was selected as an effective alternative to collect a large amount of data. For the convenience of conducting an online study, I decided to collect data for each frame of video games. In other words, study participants saw game screenshots and provided a self-report of their emotional response.

The study in this thesis had three main goals: The first goal was to replicate the supporting evidence for EnviroPulse (Valtchanov & Hancock, 2015) in the context of video games. EnviroPulse is a system to automatically determine and communicate the expected affective valence (EAV) of environments to individuals by introducing real-time affective visual feedback of the calculated EAV of images. In my work, I extend the application of this system to the area of video games and compare its predicted results with participants’ reported scores.

The second goal of this study was to analyze the emotional ratings collected to build a prediction model, for example using linear regression and machine learning. As a part of this prediction model, my aim is to include more dependent variables in the model to improve its accuracy, such as the demographic information of participants. The prediction accuracy of these two techniques, linear regression and neural networks, were compared and discussed in later chapters.

The third goal of this study was to build a tool that can predict expected affective valence as a designer is creating game visuals (e.g., as a Unity game engine plug-in) in real-time. The data collected in this study can also be used by future researchers to build their own machine learning model.

We formulate the following hypotheses:

- H1:** Participants’ reported pleasure scores are related to visual properties of game screenshots (e.g., colour and spatial frequency).
- H2:** Participants’ reported arousal scores are related to visual properties of game screenshots (e.g., colour and spatial frequency).

3.2 Participants

100 participants (52 identifying as women, 48 as men) were recruited using Prolific (<https://prolific.co/>), a third-party online survey recruitment tool. The ages of participants

ranged from 18 to 64 ($Mdn = 25$). Exclusion criteria included participants who suffer from any visual impairments, such as having a “lazy eye,” “crossed eyes,” or “color blindness.” The countries of residence of these participants are also distributed around the world, from North America and South America to Europe.

We prepared 100 unique image sets, one per participant. Participants were expected to spend about 10 seconds rating emotional reactions to each image. The whole study session took about 10 minutes for which participants received £0.87 (1.35 CAD).

3.3 Design

Participants were asked to view images and report their emotional response. All these images are screenshots from video games of different genres. Images were intentionally chosen to have different hue, saturation, luminance, and spatial frequencies, independent variables that were used for further data analysis.

1812 images in total were randomly selected and downloaded from IGDB (<https://igdb.com/>), an online video game database. All images were 1820 by 1024 pixels. Each participant was assigned 30 images, 12 of which were the same and the other 18 images were unique for everyone. In addition, there was one more “fake” video game screenshot used for an attention check which had a line of text clearly stating that participants should report an assigned value, instead of their true emotional reaction.

These 12 images (shown in Figure 3.1) which were seen by every participant were carefully selected. According to the algorithm presented by Valtchanov (2013), three of them were expected to have high arousal scores and high pleasure scores; three of them were expected to have high arousal scores and low pleasure scores; three of them were expected to have low arousal scores and high pleasure scores; three of them were expected to have low arousal scores and low pleasure scores.

Two dependent variables were measured in this study: participants’ reported pleasure scores and arousal scores.



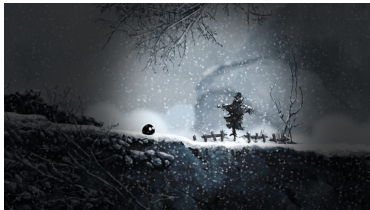
(a) Picture 1



(b) Picture 2



(c) Picture 3



(d) Picture 4



(e) Picture 5



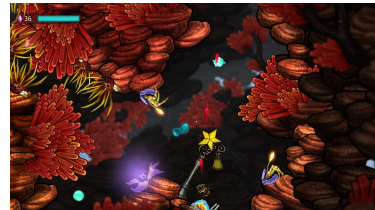
(f) Picture 6



(g) Picture 7



(h) Picture 8



(i) Picture 9



(j) Picture 10



(k) Picture 11



(l) Picture 12

Figure 3.1: Twelve images which were seen by every participant. The EAV values shown in brackets were calculated using Equation 2.1 (Valtchanov & Hancock, 2015)

3.4 Materials

Images used in this study were screenshots of different kinds of games. They were collected from IGDB (<https://www.igdb.com/>), an online video game image database. The dimensions of all images are 1820×1024 pixels. All images were saved in PNG format. When displayed during the experiment, images were scaled according to the native resolutions of participants' own devices. There were no modifications to the image content.

The whole image set can be accessed in an OSF repository¹. The graphical styles included in these images are also quite distinct. There are pixel games and modern 3D games, or possibly any other kinds of games. In addition, these images have a wide range of content, including natural scenes (shown in Figure 3.2a), city scenes (shown in Figure 3.2b), and the unreal world (shown in Figure 3.2c).



(a) A sample image which shows nature scene. (b) A sample image which shows city scene. (c) A sample image which shows unreal world.

Figure 3.2: Sample images in the image set.


3.5 Procedure

After agreeing to the consent form, participants were given instruction on how to respond using the Affective Slider (Betella & Verschure, 2016), a digital scale for the self-assessment of emotion composed of two separate slider controls (or “sliders”) that measure pleasure and arousal. More details about Affective Slider are introduced in section 2.2. Then, participants were asked to rate their emotional responses to each image by dragging the Affective Slider for 30 images (a sample survey question is shown in Figure 3.3), in addition to one attention check question, which asked the participants to read the text in the image and move the sliders to the required position. Once the images were all rated, the study


¹https://osf.io/pq8nd/?view_only=40a7012f090840f1a4443f19f9f0122e

was complete (no additional follow-up questions were asked). Participants took 8 minutes on average ($SD = 4.67\text{min}$).


Please rate the picture using **BOTH the sliders** (their order of appearance will change randomly). **Don't think too much** about it, just rate how you feel when watching it.



Move the slider to rate your level of Arousal



Move the slider to rate your level of Pleasure



→

Figure 3.3: Sample survey question.

Prolific also provided demographic information about each participant, including gender, age, country of residence, and employment status, in addition to time spent on the study session and their Prolific score (an internal rating of the quality of past participation from each participant).

3.6 Results

As long as participants completed all questions in the survey, their submissions were considered valid. There was only one participant who failed to submit a survey with valid data, and that participant's data was removed.

We analyzed the 12 images in common separately from the 1800 images rated by only one participant each. This section focuses on analyzing the 12 images in common and the 1800 images in the image pool are used for building a prediction model (discussed in chapters 4 and 5).

Figure 3.4 shows 100 participants’ arousal/pleasure ratings of these 12 images. In general, the 100 pleasure/arousal scores for each image are very scattered.

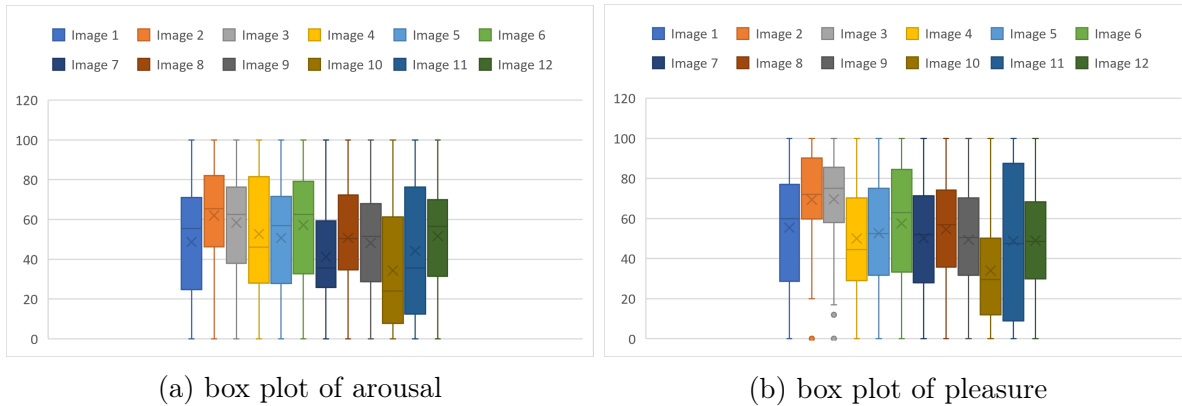


Figure 3.4: Box plots of arousal scores and pleasure scores

A one-way Analysis of variance (ANOVA) was performed to compare the effect of different images on reported arousal/pleasure scores. Table A.1a and Table A.1b shows the analysis of variance of arousal scores. Table A.2a and Table A.2b shows the analysis of variance of pleasure scores. An ANOVA revealed that there was a statistically significant difference in mean arousal score between at least two images ($F_{11,1187} = 7.24, p < 0.01$). An ANOVA also revealed that there was a statistically significant difference in mean pleasure score between at least two images ($F_{11,1188} = 11.73, p < 0.01$).

The calculation of confidence interval is shown in Figure 3.5 and Figure 3.6 for arousal scores and pleasure scores. Confidence levels are compared with expected score (calculated using Valtchanov and Hancock (2015)’s algorithm). For arousal scores, the expected score is within the 95% confidence interval for 4 of the 12 images. For pleasure scores, the expected score is within the 95% confidence interval for 2 of the 12 images. Therefore, the study data doesn’t follow the expected pattern exactly.

We also filtered the data based on whether participants passed the attention check or not. We filtered all data into three groups: the group which contains all data, the group which contains data from participants who passed the attention check, and the group which contains data from participants who failed the attention check. However, the average scores

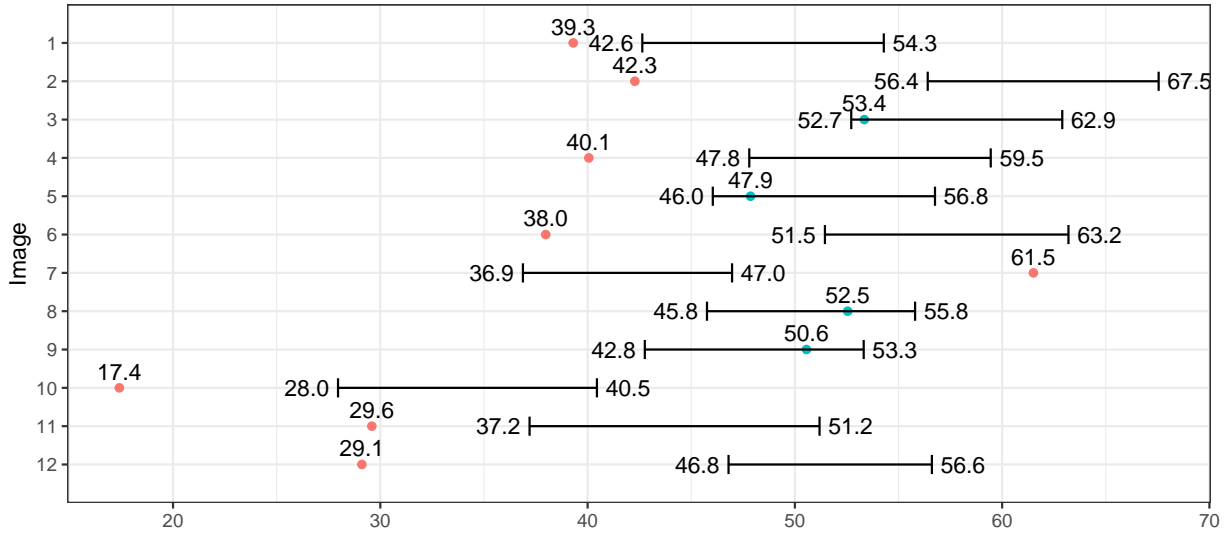


Figure 3.5: 95% confidence intervals for arousal scores for the 12 images.

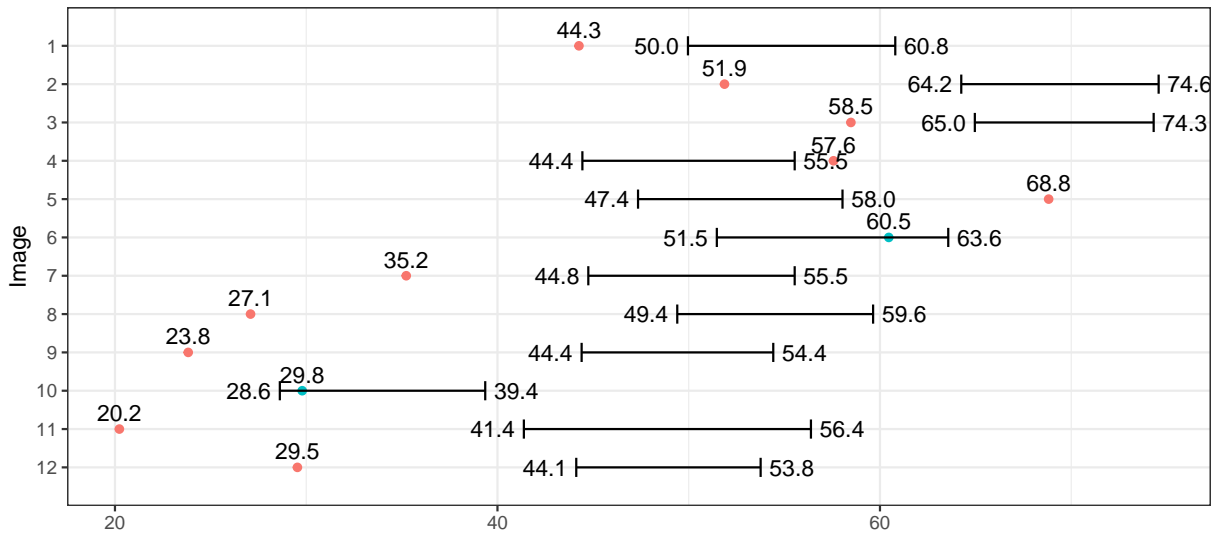


Figure 3.6: 95% confidence intervals for pleasure scores for the 12 images.

of the pass-attention-check group do not always have the smallest absolute error (compared to the expected score) among all three groups, as shown in Figure 3.7 and Figure 3.8 it appears that the average scores of participants who passed and failed the attentions check are very close. The number of times that the expected scores are within the confidence interval are quite similar between pass-attention-check data, fail-attention-check data, and entire data set (shown in Table 3.1). This lack of difference may be because participants did not notice the message within the attention check image, and so had valid responses, despite failing the attention check. As a result, we decided to use all data with out any pass-attention-check filtering when applying other techniques like linear regression and neural network to analyze the data.

	number of image that expected score is within CI (12 in total)
arousal (all)	4
arousal (pass attention check)	3
arousal (fail attention check)	3
pleasure (all)	2
pleasure (pass attention check)	0
arousal (fail attention check)	2

Table 3.1: Number of images where the expected score is within the confidence interval.

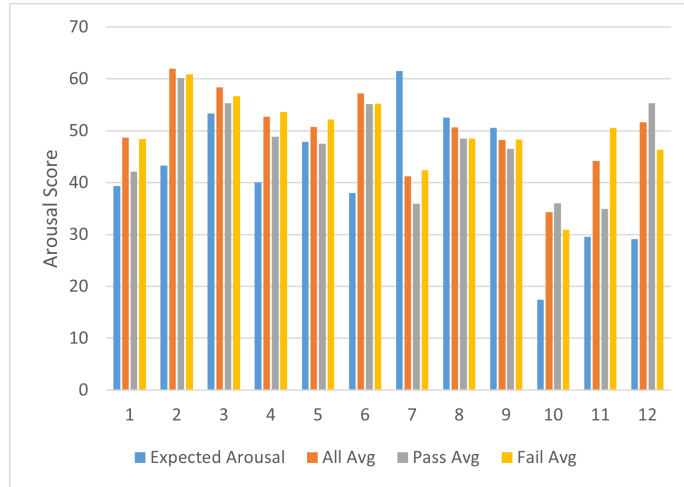


Figure 3.7: Comparison of arousal scores, from left to right: expected values calculated using Equation 2.1, mean of all participants, mean of participants that passed the attention check, mean of participants that failed the attention check.

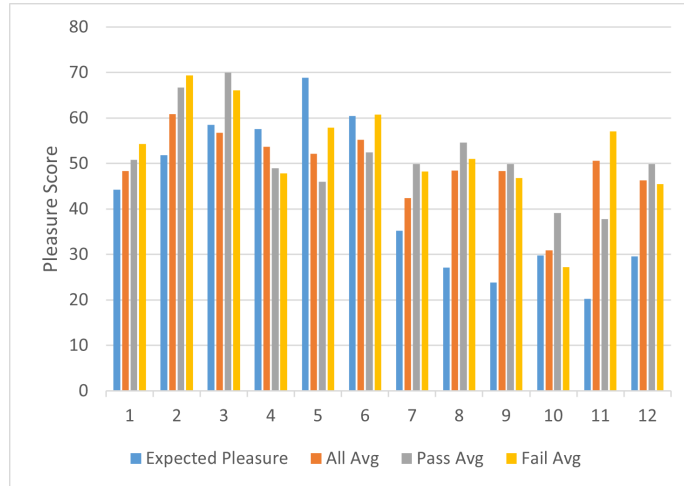


Figure 3.8: Comparison of pleasure scores, from left to right: expected values calculated using Equation 2.1, mean of all participants, mean of participants that passed the attention check, mean of participants that failed the attention check.

3.7 Summary

This chapter mainly described the online study we conducted to collect people's emotional responses to video game screenshots and analyzed the 12 images in common for each participant. ANOVAs revealed that there was a statistically significant difference in mean arousal/pleasure score between at least two images. However, the exact scores of each image do not follow the expectations exactly. Chapters 4 and 5 discuss how we used other techniques to analyze the other 1800 images.

Chapter 4

Prediction Model Using Multiple Linear Regression

4.1 Formulation of Multiple Linear Regression Model

As mentioned in previous chapters, the image pool, which has 1800 unique images only rated once by participants, were used to apply methods of linear regression and machine learning to build a prediction model. This chapter focuses on the discussion of using linear regression to analyze the data collected through the online study.

In our linear regression model, the affective levels (pleasure scores and arousal scores) are on a scale from 0 (highly negative) through 50 (neutral) to 100 (highly positive).

The raw image data cannot easily be fed into a linear regression model directly. We extracted six visual information variables as inputs of the model: colour hue, colour saturation, luminance, and the average radial power of each of the three ranges of spatial frequencies (Low range is from 0 to 10, median range is from 10 to 20, high range is above 20). The mapping from image colour and spatial frequencies to their resulting affective levels are based on the aggregation of past findings from different sources (Amir et al., 2011; Bar & Neta, 2007; Valdez & Mehrabian, 1994; Valtchanov, 2013; Valtchanov & Hancock, 2015)

To make hue, saturation, and luminance fit in a linear regression model, we first use algorithms introduced in previous research (Fernandez & Wilkins, 2008; Juricevic et al., 2010; Torralba & Oliva, 2001; Valtchanov, 2013) to map them from colour space to affective

levels. The non-linear relations between colour hue, colour saturation, colour luminance and their affective levels are show in Figure 2.2a.

Humans have different preferences for shape, complexity, and contour with the spatial frequencies of an image. Similarly, the radial power of spatial frequencies can be mapped to affective levels as well. The non-linear relations between shape, complexity, contour, and their affective levels are shown in Figure 2.2b.

Therefore, we have six explanatory predictors:

$$\begin{aligned}
 x_1 &= \text{colour hue} \\
 x_2 &= \text{colour saturation} \\
 x_3 &= \text{luminance} \\
 x_4 &= \text{radial power of high spatial frequencies} \\
 x_5 &= \text{radial power of median spatial frequencies} \\
 x_6 &= \text{radial power of low spatial frequencies}
 \end{aligned}$$

We also have two dependent variables, the arousal and pleasure scores:

$$\begin{aligned}
 y_1 &= \text{arousal score} \\
 y_2 &= \text{pleasure score}
 \end{aligned}$$

Given two data sets $\{y_{i1}, x_{i1}, \dots, x_{ip}\}_{i=1}^n$ and $\{y_{i2}, x_{i1}, \dots, x_{ip}\}_{i=1}^n$, where i is the image index from 1 to 1800, with six explanatory predictors (x_{i1} to x_{ip} , p is the independent variable index, starting from 1 to 6), and arousal score observations (y_{i1}) and pleasure score observations (y_{i2}), both the arousal model and pleasure model assume that the relationship between the vector of explanatory predictors and resulting predicted score is linear. Thus, the model has the following form:

$$\begin{aligned}
 y_{ia} &= \beta_{a0} + \beta_{a1}x_{i1} + \beta_{a2}x_{i2} + \beta_{a3}x_{i3} + \beta_{a4}x_{i4} + \beta_{a5}x_{i5} + \beta_{a6}x_{i6} + \varepsilon_{ia} \\
 y_{ip} &= \beta_{p0} + \beta_{p1}x_{i1} + \beta_{p2}x_{i2} + \beta_{p3}x_{i3} + \beta_{p4}x_{i4} + \beta_{p5}x_{i5} + \beta_{p6}x_{i6} + \varepsilon_{ip}
 \end{aligned} \tag{4.1}$$

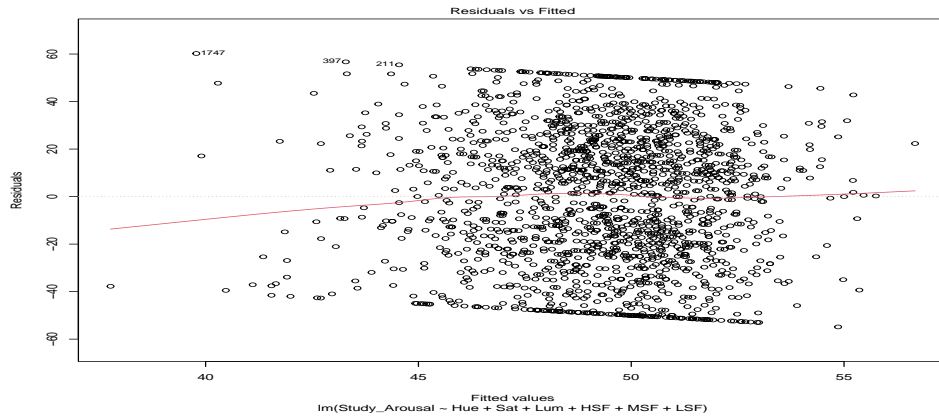
where y_{ia} are the arousal response variables for image with index i (from 1 to 1800) and y_{ip} are the pleasure response variables for image with index i (from 1 to 1800), β_{a0} and β_{p0} the intercepts, and β_{a1} to β_{a6} and β_{p1} to β_{p6} the slope coefficients for each explanatory predictor. ε_{ia} and ε_{ip} are the error terms in the linear regression model. This model can also be written in matrix form as:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{4.2}$$

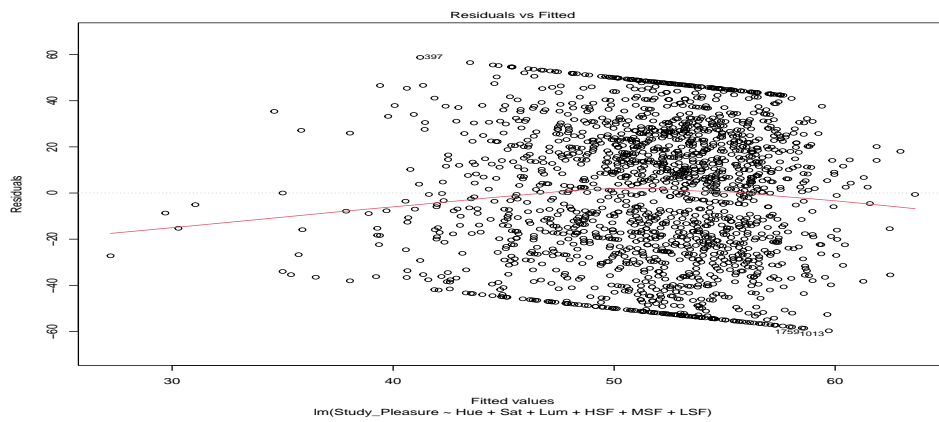
where \mathbf{Y} is a matrix of response variables, \mathbf{X} is a matrix of predictor variables with 7 columns (including 1 for the regression constant), \mathbf{B} is a matrix of regression coefficients, and \mathbf{E} is a matrix of errors.

4.2 Examination of Linearity Assumption

The linear regression model to predict arousal score and pleasure score were calculated using R studio (shown in Listing A.1 and Listing A.2 separately). When examining the model, we can validate the linearity assumption. The assumptions of multivariate regression analysis are normal distribution, linearity, freedom from extreme values, and having no multiple ties between independent variables. The linearity assumption of the arousal/pleasure prediction model is demonstrated in figs. 4.1 to 4.4. Note that the red lines in the figures are locally weighted scatterplot smoothing (LOWESS), which smooth over the scattered points to look for certain kinds of patterns in the y-axis values (Cleveland, 1981).

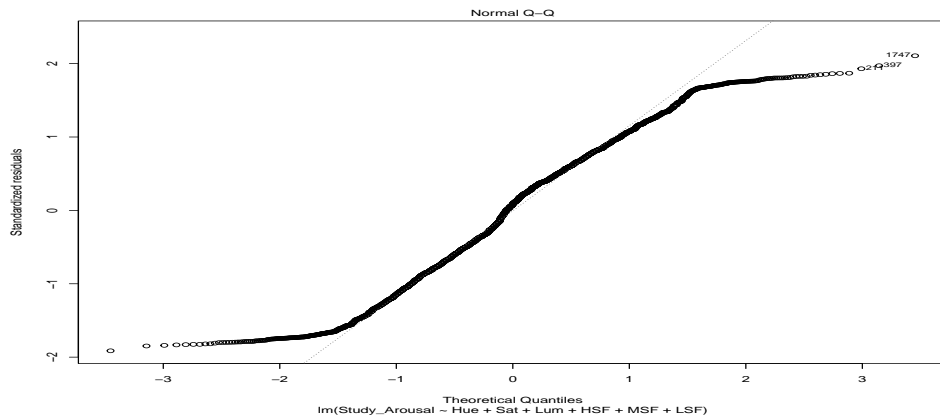


(a) Residuals vs Fitted plot of the arousal model

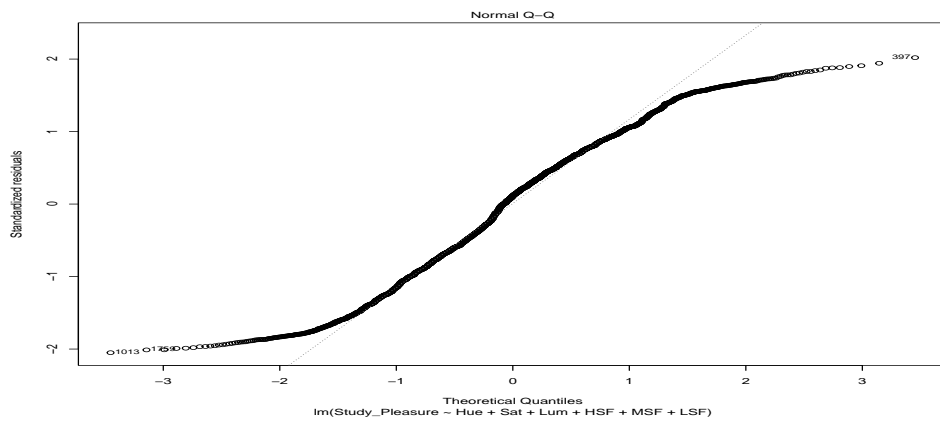


(b) Residuals vs Fitted plot of the pleasure model

Figure 4.1: Residuals vs Fitted plot of the arousal/pleasure model. The LOWESS (red line) in the graphs of residuals vs. fitted data is relatively flat, which provides evidence for the linearity assumption.

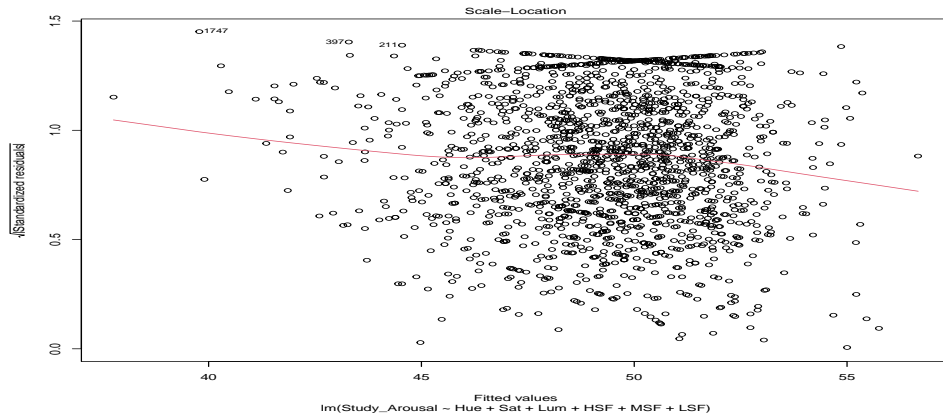


(a) Normal Q-Q plot of the arousal model

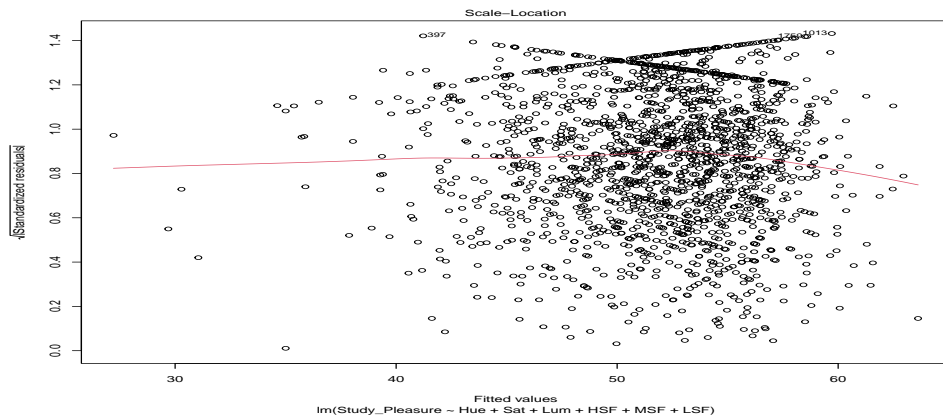


(b) Normal Q-Q plot of the pleasure model

Figure 4.2: Normal Q-Q plot of the arousal/pleasure model. The points form a roughly straight line that follows the diagonal. This means the residual errors are roughly normally distributed and provides evidence for the linearity assumption.

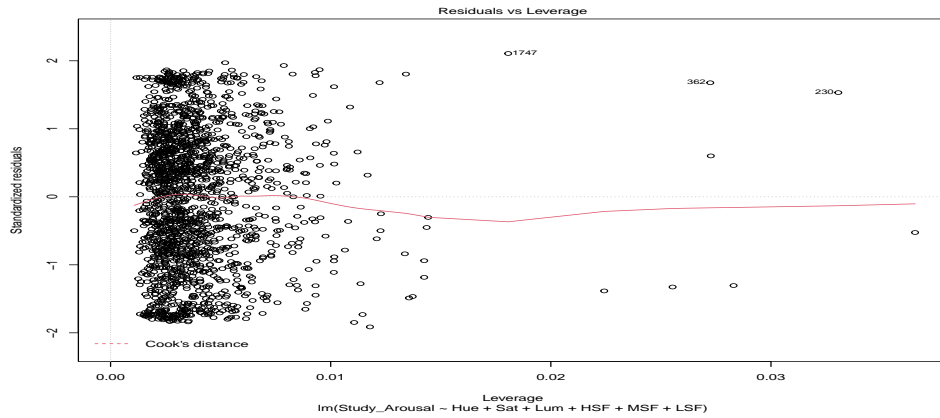


(a) Scale-Location plot of the arousal model

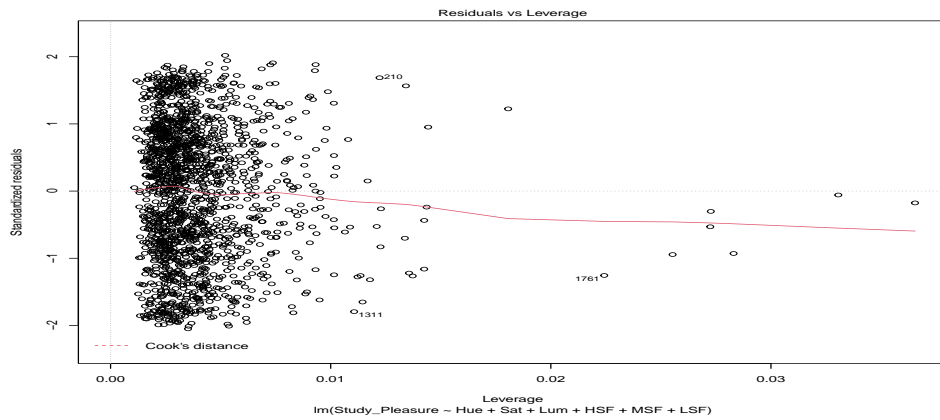


(b) Scale-Location plot of the pleasure model

Figure 4.3: Scale-Location plot of the arousal/pleasure model. The LOWESS (red line) in the graphs of Scale-Location is relatively flat, which provides evidence for the linearity assumption.



(a) Residuals vs Leverage plot of the arousal model



(b) Residuals vs Leverage plot of the pleasure model

Figure 4.4: Residuals vs Leverage plot of the arousal/pleasure model. The spread of standardized residuals in both graphs do not change as a function of leverage (the LOWESS (red line) is relatively flat), which provides evidence for the linearity assumption.

4.3 Multiple Linear Regression Results

A multiple linear regression was calculated to predict arousal based on hue, saturation, luminance, and high, medium, and low, spatial frequency. This regression equation was not significant ($F_{6,1793} = 1.98$, $p = .066$, $R^2 = .007$).

$$\begin{aligned} y_{\text{arousal}} = & 30.342 \\ & +0.004 \text{ (hue's affective level)} \\ & +0.016 \text{ (saturation's affective level)} \\ & +0.032 \text{ (luminance's affective level)} \\ & -0.215 \text{ (HSF's affective level)} \\ & +0.158 \text{ (MSF's affective level)} \\ & +0.095 \text{ (LSF's affective level)} \end{aligned} \tag{4.3}$$

Only high spatial frequency ($\beta = -0.215$, $p = .004$) was a significant predictor of arousal. No other predictors were significant ($p > .1$).

The same multiple linear regression was calculated to predict pleasure. A significant regression equation was found ($F_{6,1793} = 7.06$, $p < .001$, $R^2 = .023$).

$$\begin{aligned} y_{\text{pleasure}} = & 17.595 \\ & -0.011 \text{ (hue's affective level)} \\ & +0.038 \text{ (saturation's affective level)} \\ & +0.127 \text{ (luminance's affective level)} \\ & -0.334 \text{ (HSF's affective level)} \\ & +0.333 \text{ (MSF's affective level)} \\ & +0.085 \text{ (LSF's affective level)} \end{aligned} \tag{4.4}$$

Luminance ($\beta = 0.127$, $p = .005$), high spatial frequency ($\beta = -0.334$, $p < .001$) and medium spatial frequency ($\beta = 0.085$, $p = .004$) were significant predictors of pleasure. No other predictors were significant ($p > .1$).

4.4 Examination of Independent Variable's Relative Importance

Listing A.1 and Listing A.2 show the detailed results of the multiple linear regression in R studio. R-squared (R^2) is a statistical measure that represents the proportion of the

variance for a dependent variable that is explained by an independent variable or variables in a regression model (Wright, 1921). The arousal model’s R^2 value is 0.66% and the pleasure model’s R^2 value is 2.3%. We also filtered the data based on whether participants passed the attention check. However, there was no obvious improvement on R^2 value for data that only includes participants who passed the attention check.

Both models’ R^2 values are very low. Such low R^2 value is understandable because unpredictable human behaviours and many other factors such as participant’s preference on image’s content may contribute more to the total variance than our linear regression model. It is more meaningful to examine the relative importance of each input of the model.

The relative importance of the independent variables of the linear model (shown in Table 4.1) was calculated by the variance decomposition metric LMG (named after the authors Richard H. Lindeman, Peter F. Merenda, and Ruth Z. Gold), which is simply R^2 partitioned by averaging over orders (Lindeman et al., 1981).

	Relative importance in arousal model	Relative importance in pleasure model
Hue	0.78%	0.70%
Sat	5.49%	7.42%
Lum	9.10%	26.36%
HSF	34.28%	19.05%
MSF	20.71%	24.82%
LSF	29.65%	21.65%

Table 4.1: Independent variable’s relative importance in arousal/pleasure model

15.37% of the total variance in the arousal model comes from independent variables in the colour space, while 34.48% from colour space for pleasure model. Compared with independent variables in the colour space, the spatial frequencies have a much larger impact on the predicted arousal score and pleasure score. This is in accordance with previous research, stating that the colour score (hue, saturation and luminance’s affective level) and content score (spatial frequency’s affective level) can be combined into the EAV score using weights inferred from the variance and regression analysis by Valdez and Mehrabian (1994), and by Valtchanov (2013), as follows (more details in section 2.5):

$$EAV \text{ Score} = 0.3 \cdot \text{colour score} + 0.7 \cdot \text{content score} \quad (4.5)$$

4.5 Comparison of Multiple Linear Regression Model and Valtchanov’s algorithm

Both Valtchanov (2013)’s algorithm and the multiple linear regression model were applied to the same 1800 images used in the online study to predict arousal/pleasure. Table 4.2 shows the comparison between the multiple linear regression model and Valtchanov (2013)’s algorithm, including the L1 loss, R^2 , Pearson correlation coefficient (PCC) between predicted scores and actual scores, and the predicted score’s standard deviation (SD). The linear regression model performs much better than Valtchanov (2013)’s algorithm in terms of L1 loss and PCC. However, its SD is extremely low (below 5), which means that predicted scores are clustered around the mean and the prediction is not meaningful in this way. Also, the MLR’s R^2 is really low, which means the our MLR models cannot explain the variance of study data well.

	L1 Loss	R^2	PCC	SD
Valtchanov (2013) arousal score	30.060	0.001	0.070	8.510
Valtchanov (2013) pleasure score	30.090	0.001	0.030	14.613
MLR arousal score	24.630	0.007	0.140	2.338
MLR pleasure score	24.600	0.024	0.150	4.612

Table 4.2: Comparison of Multiple Linear Regression Model and Valtchanov (2013)’s algorithm

4.6 Summary

This chapter examines the linearity assumption of the collected data and used a linear regression model to predict arousal/pleasure. However, it was found that multiple linear regression is not an ideal solution to build a prediction model due to its low R^2 and low SD. We explored using a neural network instead in Chapter 5.

Chapter 5

Neural Network Models to Predict Affect

As discussed in Chapter 4, the multiple linear regression model is not an ideal solution to build a prediction model using the study data, since its R^2 is extremely low, suggesting that it cannot explain the variance in the study data very well. In this chapter, I explore the use of a neural network. This chapter discusses three different neural network models. Their performance was compared with the multiple linear regression model from Chapter 4 and Valtchanov and Hancock (2015)'s algorithm.

5.1 Training & Dataset

Our neural network models were trained using an online platform called Weights & Biases¹, which helps developers to automate hyper-parameter optimization and explore the space of possible models. In this research, Weights & Biases was mainly used to tune the learning rate, the batch size and the number of neurons in the network layers. The goal of the training process is to achieve convergence of the loss function and minimize the converged loss.

The dataset used to train our neural network model is the same dataset used in the linear regression model in Chapter 4, which includes 1800 unique images. The method of K-fold cross-validation is used to re-sample the dataset and test the machine learning

¹<https://wandb.ai>

model’s performance (Stone, 1976). In each training epoch, the whole dataset is split into three portions: 80% of the data set is the training set, 10% of the data set is the validation set, and the last 10% of the data set is the test set.

5.2 Multi-layer perception network

A multi-layer perception (MLP) network is a basic type of feed-forward artificial neural network (ANN). It has three components: an input layer, a hidden layer and an output layer. MLP networks are sometimes called traditional neural networks. They are different from linear regression models in their multilayer structure and nonlinear data relationships. All the nodes in the network are called neurons. The transmission of neurons in the network is controlled by nonlinear activation functions (Hastie et al., 2009). In this thesis, a rectified linear unit (ReLU) function is used for the activation function in all neurons using Equation 5.1. Compared to a logistic sigmoid function and a hyperbolic tan function, ReLU performs better. It can solve the problems of gradient explosion and gradient disappearance and maintains its convergence rate in a stable state (Zoph & Le, 2018).

$$\text{ReLU}(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \quad (5.1)$$

Since the network is fully connected, each node in one layer connects to every node in the next layer with a certain weight. Therefore, the output c of each neuron is:

$$c = \varphi \left(\sum w_i a_i + b \right) \quad (5.2)$$

where a_i and w_i are the inputs and weights of the neuron respectively, b is the bias of the current neuron and φ is the activation function ReLU.

In this thesis, the MLP network is composed of three parts (an input layer, multiple hidden layers and an output layer) of nonlinearly-activating nodes (Figure 5.1a). The input layer has 6 explanatory predictors (hue/saturation/luminance/HSF/MSF/LSF’s affective score), the same as the linear regression model discussed in Chapter 4. The detailed structures of the hidden layers are shown in Figure 5.1b. All of the weights and biases of the linkages between the potential variables and predicted affective values are updated when the parameters of the network are learned based on the training data. Then the predicted values are calculated based on the learned weights and biases when the input data is applied. The output layer has only one neuron, which represents the values of the arousal or pleasure scores.

Figures A.1 to A.3 show the loss during the training process of the MLP network to predict the pleasure score. The loss displayed here is $L1$ loss, also known as absolute error loss, which is the absolute difference between a prediction and the actual value. An exponential moving average is implemented to the loss plot (Gardner, 1985) in the range 0 to 1 with the smoothing value of all loss plots of 0.5.

The plot of training loss (Figure A.1) decreases to a point of stability. The plot of validation loss (Figure A.2) also decreases to a point of stability. The learning curves show no sign of over-fitting or under-fitting. For pleasure's MLP model, the smoothed training loss is 24.957, the smoothed validation loss is 24.86 and the smoothed test loss is 25.815.

Figures A.4 to A.6 show the loss during the training process of the MLP network to predict the arousal score. Similar to the training process of pleasure model, the plot of training loss (Figure A.4) decreases to a point of stability. The plot of validation loss (Figure A.5) also decreases to a point of stability. The learning curves show no sign of over-fitting or under-fitting. For arousal's MLP model, the smoothed training loss is 26.071, the smoothed validation loss is 24.865 and the smoothed test loss is 24.913.

The loss of the simple MLP network are all relatively high as the next section shows that it can be further decreased by incorporating a pre-trained residual neural network (ResNet)-50 model.

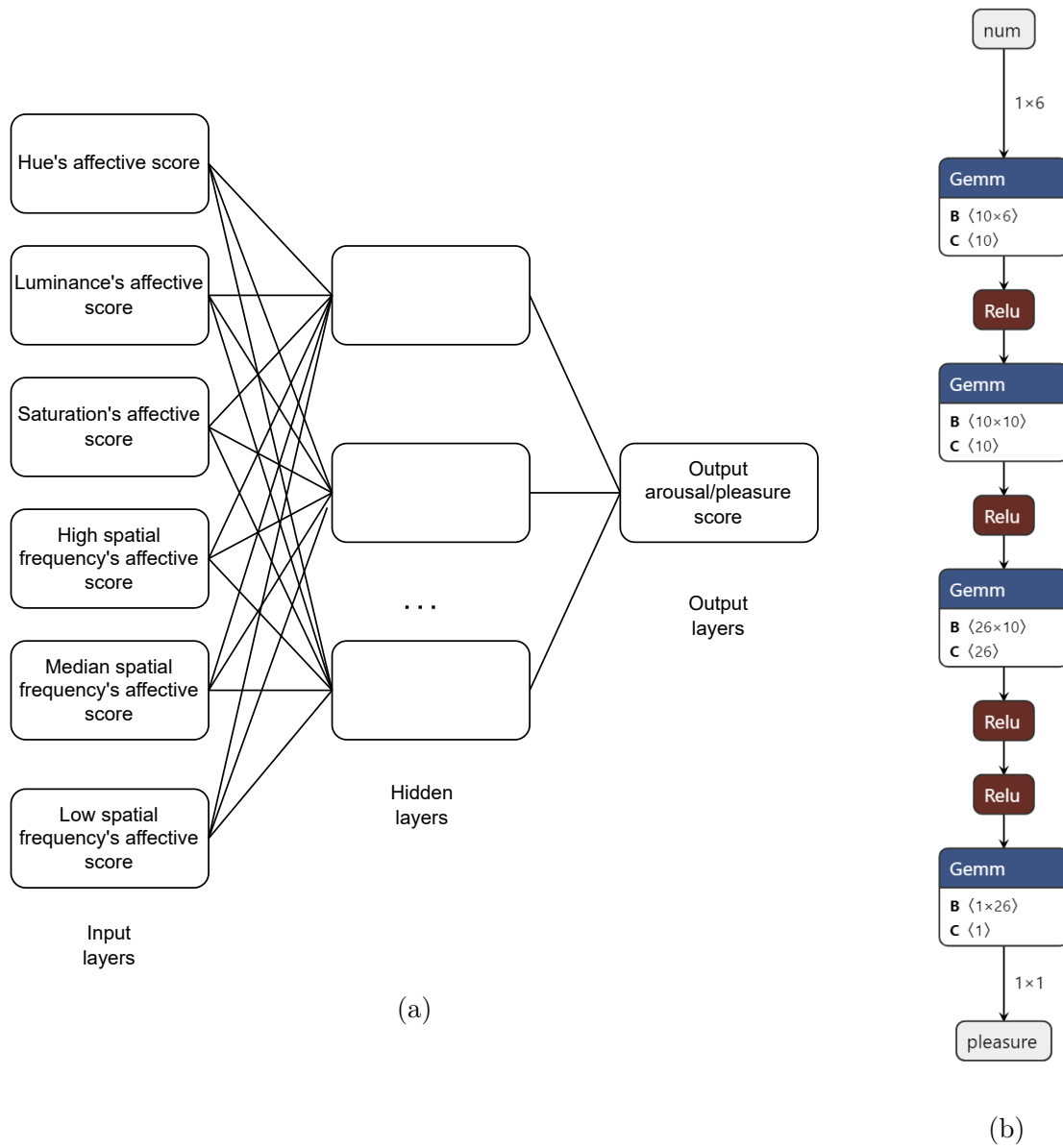


Figure 5.1: Structure of the MLP network for predicting arousal/pleasure score: (a) shows the general structure and (b) shows the details of hidden layers, including numbers of inputs/outputs and activation functions.

5.3 Convolutional neural network with ResNet-50

Transfer learning is a technique where knowledge gathered from one task/model is used in another similar task (Bozinovski, 2020). As introduced in section 2.2, a ResNet is an innovative neural network that was first introduced by He et al. in 2016. ResNet has many variants that run on the same concept but have different numbers of layers. ResNet-50 is used to denote the variant that can work with 50 neural network layers.

We included a pre-trained ResNet-50 model into our MLP network, in the hopes of increasing the accuracy of the affective score prediction. ResNet-50 can detect many low, mid, and high-level features in the image, which may all contribute to the arousal and pleasure scores. The new network combines the output data of the ResNet-50 with the output data of the MLP. As shown in Figure 5.2, the output layer of the MLP concatenates with the output layer of the ResNet-50, forming a fully-connected layer, which then outputs the arousal score or pleasure score.

Figures A.7 to A.9 show the $L1$ loss during the training process of the MLP combined with ResNet-50 to predict the pleasure score. The plot of training loss (Figure A.7) decreases to a point of stability. The plot of validation loss (Figure A.8) also decreases to a point of stability. The learning curves show no sign of over-fitting or under-fitting. The smoothed training loss (24.411), the smoothed validation loss (23.583) and the smoothed test loss (23.761) of the MLP combined with ResNet-50 are less than those of the simple MLP.

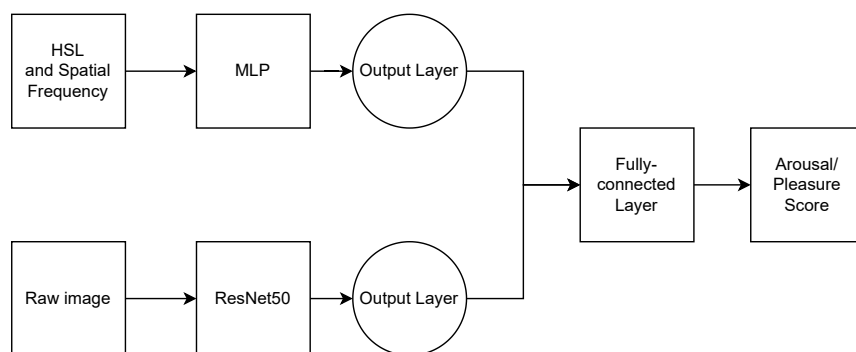


Figure 5.2: Structure of the MLP network combined with ResNet-50.

Figures A.10 to A.12 show the $L1$ loss during the training process of the MLP combined with ResNet-50 to predict arousal score. The plot of training loss (Figure A.10) decreases to a point of stability. The plot of validation loss (Figure A.11) also decreases to a point of

stability. The learning curves show no sign of over-fitting or under-fitting. The smoothed training loss (22.412), the smoothed validation loss (23.646) and the smoothed test loss (23.886) of the MLP combined with ResNet-50 are less than those of the simple MLP.

5.4 Integration of user-specific profiles

The online survey platform, Prolific, provides detailed user information including sex, age, and nationality. As mentioned in section 2.3, previous research found that men and women have different preferences when looking at low- and medium-wavelength colours. Women prefer redder colours while men prefer colours that are more blue-green (Palmer & Schloss, 2010). We explored using these user-specific profiles to improve the accuracy of our neural network model. Thus, one numerical data type (age) and two categorical data (sex and country of residence) are included in our model’s input data.

The technique of categorical embedding has been applied to the make the categorical data fit in the neural network model. An embedding layer is created for each of the categorical features. This approach allows for a relationship between categories to be captured by converting raw categories into embeddings and concatenating them together with the rest of the other numerical features (Dahouda & Joe, 2021). The structure of the new model is shown in Figure 5.3. The goal is that the categorical embeddings can capture more rich/complex relationships that will ultimately improve the performance of the model.

Figures A.13 to A.15 show the $L1$ loss during the training process of the MLP combined with ResNet-50 and user-profile integration to predict the pleasure score. The plot of training loss (Figure A.13) decreases to a point of stability. The plot of validation loss (Figure A.14) also decreases to a point of stability. The learning curves show no sign of over-fitting or under-fitting. The smoothed training loss (23.61), the smoothed validation loss (23.323) and the smoothed test loss (23.117) of the MLP combined with ResNet-50 are less than those of the MLP combined with ResNet-50 only.

Figures A.16 to A.18 show the $L1$ loss during the training process of the MLP combined with ResNet-50 and user-profile integration to predict arousal score. The plot of training loss (Figure A.16) decreases to a point of stability. The plot of validation loss (Figure A.17) also decreases to a point of stability. The learning curves show no sign of over-fitting or under-fitting. The smoothed training loss (21.893), the smoothed validation loss (21.707) and the smoothed test loss (22.679) of the MLP combined with ResNet-50 are less than those of the MLP combined with ResNet-50 only.

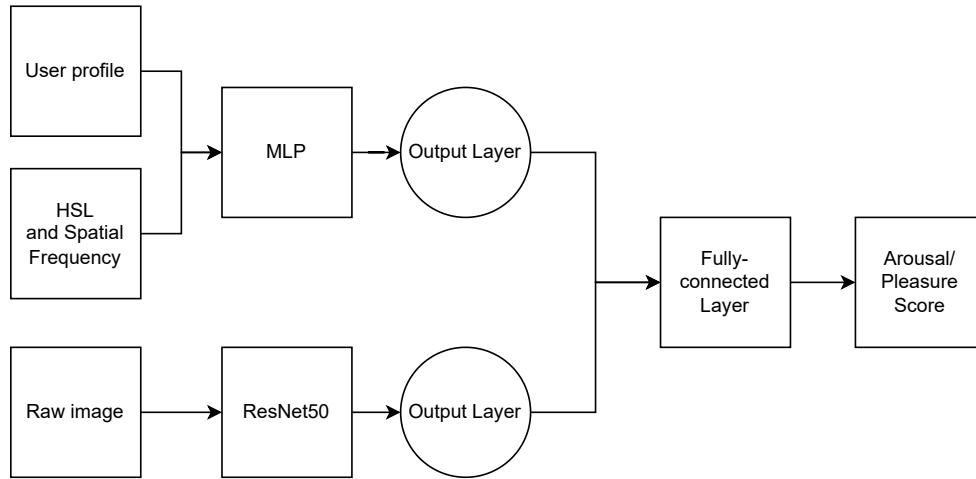


Figure 5.3: Structure of MLP network combined with ResNet-50 and user-profile integration.

5.5 Performance comparison

Table 5.1 shows the comparison of $L1$ loss among three different neural network models introduced in sections 5.2 to 5.4. In order to compare these methods, an error indicator ($L1$ loss) and a fitness indicator (R^2) were applied into testing the whole prediction results (Naser & Alavi, 2021). A low $L1$ loss and a high R^2 indicate good performance. As can be seen from this table, the $L1$ loss decreases as the model becomes more complex.

When applying Valtchanov and Hancock (2015)’s algorithm to the same data set (1800 images from the study) to predict pleasure/arousal, the average $L1$ loss is 30.06 for arousal and 30.09 for pleasure, both of which are much larger than all neural network models introduced in this chapter.

	Smoothed training loss	Smoothed validation loss	Smoothed test loss
MLP	24.957	24.86	25.815
MLP + ResNet-50	24.411	23.583	23.761
MLP + ResNet-50 + user-profile	23.61	23.323	23.117

Table 5.1: Comparison of different neural network models to predict pleasure score

	Smoothed training loss	Smoothed validation loss	Smoothed test loss
MLP	26.071	24.865	24.913
MLP + ResNet-50	22.412	23.646	23.886
MLP + ResNet-50 + user-profile	21.893	21.707	22.679

Table 5.2: Comparison of different neural network models to predict arousal score

Table 5.3 compares the performance of all five prediction models introduced in this thesis. Valtchanov and Hancock (2015)’s algorithm, multiple linear regression, and the three neural network models were then compared for the whole data set (1800 images) being used for training and testing the models.

	Pleasure		Arousal	
	$L1$ Loss	R^2	$L1$ Loss	R^2
Valtchanov and Hancock (2015)’s algorithm	30.090	0.001	30.060	0.001
Multiple Linear Regression	24.600	0.024	24.630	0.007
MLP (Neural Net)	26.405	0.000	25.407	0.000
MLP + ResNet-50	24.878	0.049	24.825	0.032
MLP + ResNet-50 + user-profile	24.380	0.050	24.753	0.034

Table 5.3: Comparison of Valtchanov and Hancock (2015)’s algorithm, the multiple linear regression model, and the neural network models to predict pleasure and arousal scores

For all the five methods, $L1$ loss of predicting pleasure scores for the total data set ranged from 24.38 to 30.09 and $L1$ loss of predicting arousal scores ranged from 24.753 to 30.06. R^2 of predicting pleasure scores for the total data set ranged from 0 to 0.05 and R^2 of predicting arousal scores ranged from 0 to 0.034. These data show the power of the neural network to improve the prediction accuracy to some degree. Among all the five methods, the $L1$ loss of MLP with ResNet-50 and user-profile is the lowest and the R^2 is the highest.

Valtchanov and Hancock (2015)’s algorithm has the highest $L1$ loss (over 30) and really low R^2 (0.001 for both pleasure and arousal predictions), which were much lower than in the original work. This low performance might be due to several factors. First of all, in Valtchanov’s original thesis, where he conducted experiments to develop coefficients to be used in Valtchanov and Hancock (2015)’s algorithms, participants saw images of natural and urban scenes, instead of video game screenshots. Second, Valtchanov (2013)’s

experiments were all in-person and all images were displayed on a 17 inch LCD monitor with a resolution of 1280 x 1024 pixels. Participants were required to be seated at the same distance from the monitor on a chair that could not be moved. Unlike Valtchanov (2013)'s experiment setup, the online study conducted in this thesis could be viewed on any electronic device with different screen resolutions and distances to the viewers, and there was no easy way to capture this information.

The simple MLP, introduced in section 5.2, also has poor performance with relatively large $L1$ loss (26.405 for pleasure and 25.407 for arousal) and extremely low R^2 (near 0). This demonstrates that our simple MLP cannot predict arousal/pleasure score well, due to its simple structure and limited number of hidden layers. However, this may be due to similar limitations mentioned above (i.e., lack of experimental control) and it does improve upon Valtchanov and Hancock (2015)'s algorithm slightly.

A newly trained multiple linear regression model, introduced in section 4.1, performs slightly better than Valtchanov and Hancock (2015)'s algorithm and the MLP neural net with relatively low $L1$ loss (24.6 for pleasure and 24.63 for arousal). However, its R^2 is still considered very low (0.024 for pleasure and 0.007 for arousal).

As the structure of neural network becomes more complex, its performance becomes better. When MLP concatenates with ResNet-50, introduced in section 5.3, the performance was much better, shown by the larger R^2 . After including a pre-trained ResNet-50 model, R^2 increased from 0 to 0.049 for pleasure and from 0 to 0.032 for arousal. The $L1$ loss also decreased from 26.405 to 24.878 for pleasure and from 25.407 to 24.825 for arousal. Compared to the slight improvement on $L1$ loss, the increment of R^2 is quite significant.

With the user profile integrated into the neural network, the performance became even better. $L1$ loss also decreased from 24.878 to 24.38 for pleasure and from 24.825 to 24.753 for arousal. The R^2 increased very slightly from 0.049 to 0.05 for pleasure and from 0.032 to 0.034 for arousal. The model with user-specific profiles has a slightly lower $L1$ loss and a higher R^2 , which does not provide strong evidence that age, sex, and nationality affect participants' reported affective levels, though further research is needed.

It can also be shown from the comparison in Table 5.3 that the pleasure model performs better than the arousal model, when examining the methods of multiple linear regression and neural networks. The difference in $L1$ loss is quite small; however, in terms of R^2 , the multiple linear regression, MLP with ResNet-50, and MLP with ResNet-50 and user-profile all perform better when predicting pleasure. This demonstrates that the existing input variables in our methods explain the variance of reported pleasure better. Further study is required to increase the accuracy of arousal.

5.6 Summary

The chapter discusses how we built neural network models to predict pleasure and arousal based on images rendered in video games and compared their performances with multiple linear regression and Valtchanov and Hancock (2015)'s algorithm. The next chapter introduces how we incorporated these neural network models inside the Unity game engine and built a prediction tool.

Chapter 6

Unity Plug-in to Display Real-Time Predicted Affective Status

In this chapter I present a Unity game engine plug-in based on the prediction models discussed in Chapters 4 and 5. This tool derives from research on the psychological impact of images to support a novel kind of affective computing for video games. This tool automatically calculates an expected affective valence (EAV) score for the current game scene through visual analysis of image input in real-time. It can be used to guide game designers to build game scenes that promote better psychological well-being or achieve certain desired emotional feedback from game players.

6.1 How to Use

This plug-in is composed of two parts, an on-screen display indicator (Figure 6.1) and a configuration window (Figure 6.2). The on-screen display uses the Affective Slider (AS), introduced in section 2.2. The AS was originally designed as self-assessment tool to report affective status. Here in our Unity plug-in, it has the same visual format but only works



Figure 6.1: The affective slider of the Unity plugin

as feedback to indicate the current EAV and cannot be adjusted by the user. The on-screen-display is located at the bottom centre of the screen and has two parts, a pleasure indicator on the left side and an arousal indicator on the right side.

The Unity game engine can be used in two modes, Play mode and Edit mode. In the Play mode, users can view and play the game in the Game View Window. In the Edit mode, users can only edit the game scene. This plug-in allows user to evaluate player's expected affective status in both modes.

The on-screen display of this plug-in is designed to be used in Play mode. When the game starts to run inside the Unity game engine, this plug-in's on-screen display also shows up in the same window and evaluates the current rendered game scene. It displays the current expected affective values in real-time.

Still in Play mode, the configuration window of this plug-in shows more details about the affective contribution from each factor of the rendered game scene, including hue, luminance, saturation and spatial frequencies. The configuration window can also be used in Edit mode. It can continuously provide predicted affective status as game developers/designers change the game scene in the Scene view window of the Unity game engine. In Edit mode, instead of taking the rendered scene of a running game as input, this Unity plug-in takes a screenshot of the Scene view window as input. While in Edit mode, the image being tested does not correspond directly to what a player might experience, it does provide a faster way for a designer to test different scene elements, and the designer can still adjust their view to maximize the benefit of the predicted score (i.e., by approximating the player's view).

Users can also choose to load their own linear regression model or a pre-trained neural network model. They can also load a static static image from a local device and see its EAV by clicking the "Open picture" button at the bottom of the configuration window.

Users can start the real-time EAV prediction by clicking the start button in the configuration window. Figures 6.3 and 6.4 provide two examples of using the Unity plug-in. It can be seen in the figures that the visual feedback shows different values in these different game scenes. In Figure 6.3 the blue sky has a similar pleasure score to Figure 6.4 which shows green grass, but Figure 6.3 has a relatively lower arousal score. These two examples indicate that blue sky can make players feel calmer than green grass in the game.

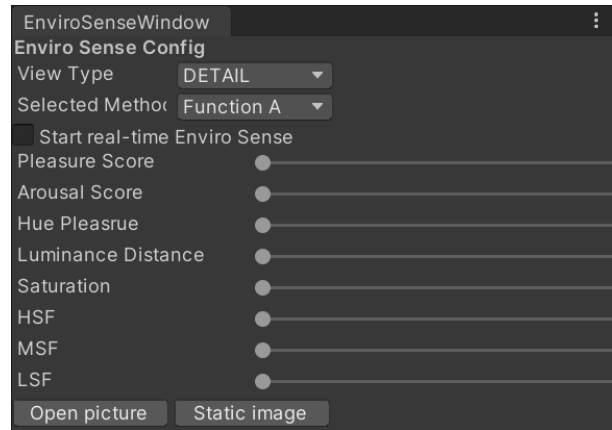


Figure 6.2: Configuration window of the Unity plugin

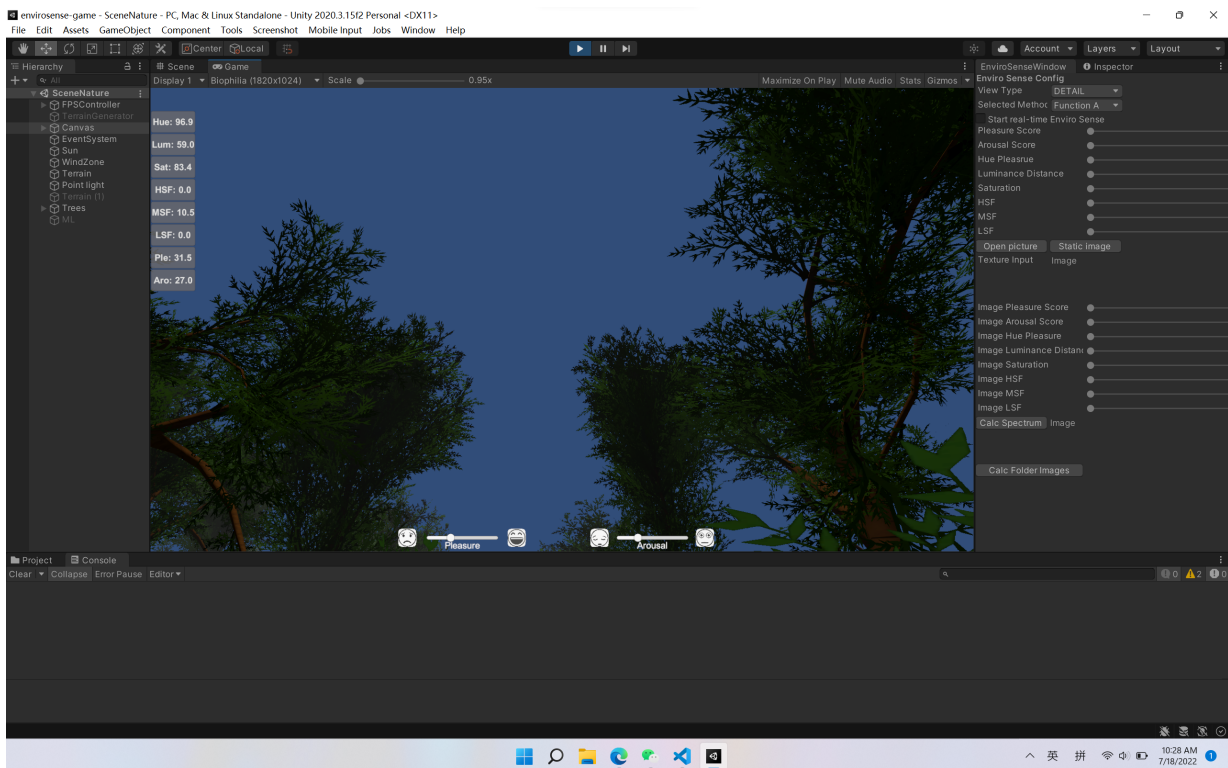


Figure 6.3: Screenshot of using the Unity plug-in (a)

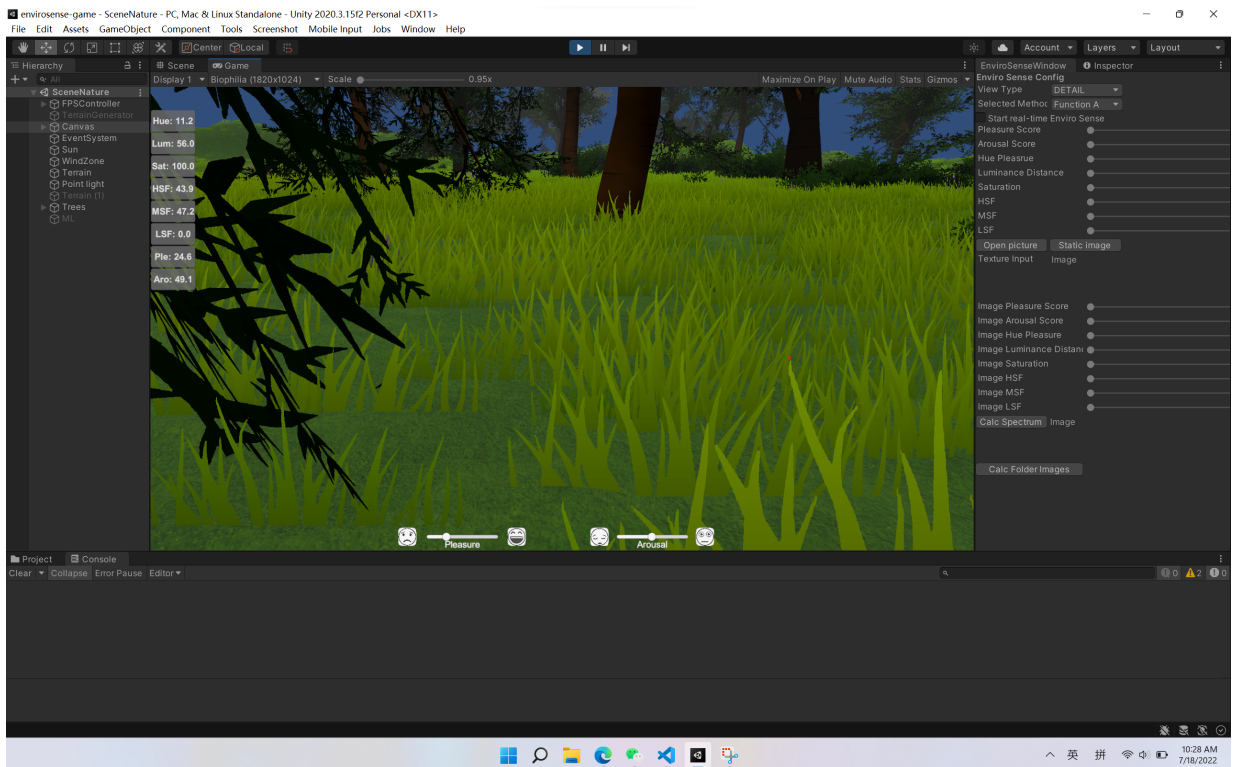


Figure 6.4: Screenshot of using the Unity plug-in (b)

6.2 Engineering Pipeline

In this section, I elaborate on the algorithms used to engineer this plug-in. Figure 6.5 shows the engineering pipeline of this Unity plug-in. It captures a screenshot of the current game scene as input and passes it to two different sub-pipelines, one in colour space and the other in spatial frequency space. The following subsections describe the engineering pipeline of this Unity plug-in. Figure 6.5 explains how the raw image data is processed and fed into the neural network.

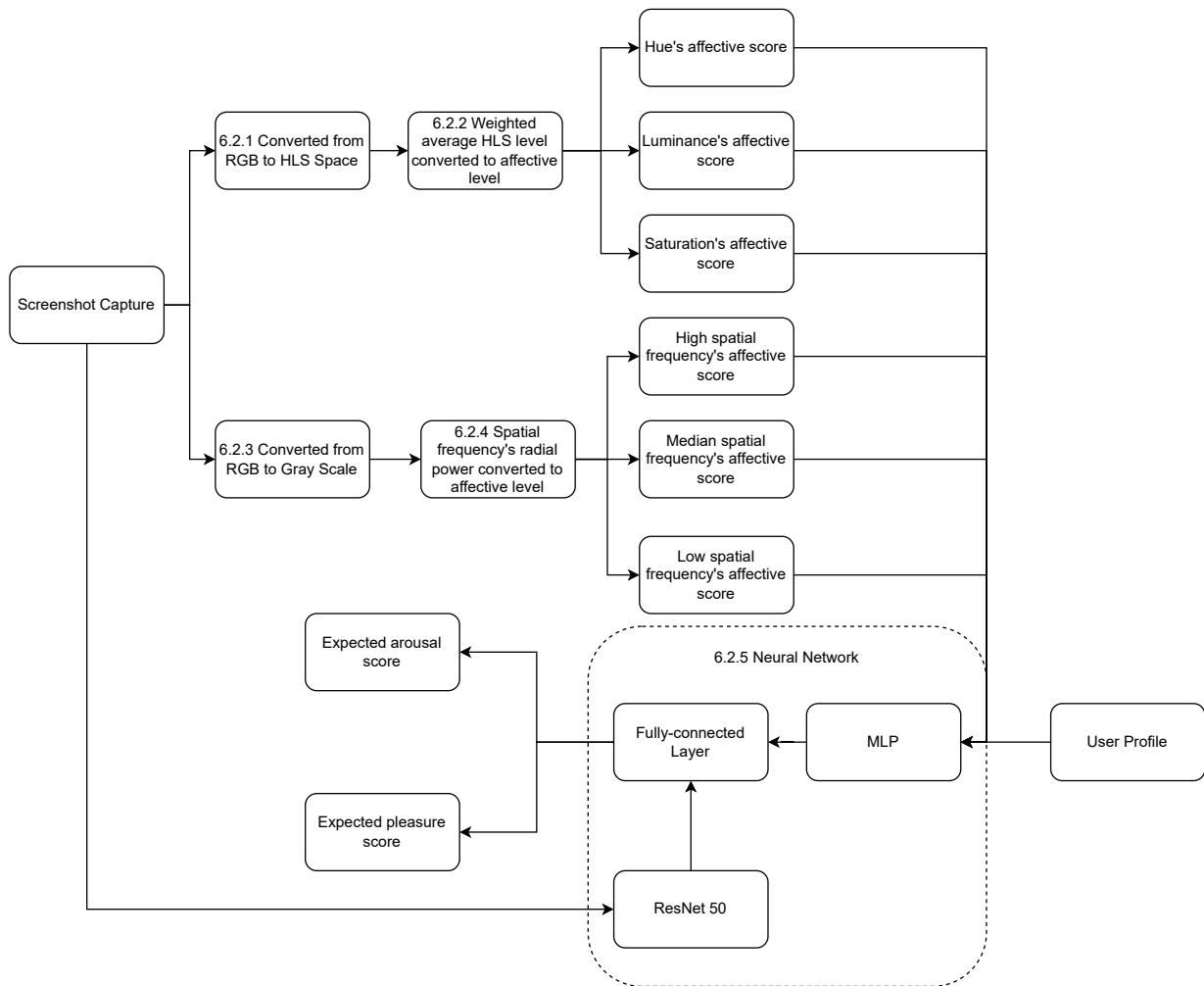


Figure 6.5: Engineering pipeline of the Unity plugin (details are explained in the indicated subsections).

6.2.1 Colour Conversion

In the sub-pipeline in colour space, the image is first converted from RGB to HSL (hue, saturation, and luminance). This conversion of the image is essential for later histogram calculation. The conversion formula is as follows (Equation 6.1):

$$\begin{aligned}
 V_{\max} &\leftarrow \max(R, G, B) \\
 V_{\min} &\leftarrow \min(R, G, B) \\
 L &\leftarrow \frac{V_{\max} + V_{\min}}{2} \\
 S &\leftarrow \begin{cases} \frac{V_{\max} V_{\min}}{V_{\max} V_{\min}} & \text{if } L < 0.5 \\ \frac{V_{\max} V_{\min}}{2 - (V_{\max} V_{\min})} & \text{if } L \geq 0.5 \end{cases} \\
 H &\leftarrow \begin{cases} 60(G - B) / (V_{\max} - V_{\min}) & \text{if } V_{\max} = R \\ 120 + 60(B - R) / (V_{\max} - V_{\min}) & \text{if } V_{\max} = G \\ 240 + 60(R - G) / (V_{\max} - V_{\min}) & \text{if } V_{\max} = B \\ 0 & \text{if } R = G = B \end{cases}
 \end{aligned} \tag{6.1}$$

6.2.2 Average Weighted Histogram

We use the OpenCV library to calculate histograms of the rendered game scene for each channel, including hue, saturation, and luminance (Bradski, 2000). The pixel counts (the values in the histogram) are multiplied by the pixel values (hue/saturation/luminance) and summed together to get a weighted sum. Then the weighted sum is divided by the number of pixels to get the weighted average of hue, saturation, and luminance across the entire image. These weighted averages are passed into next stage. The colour pipeline has three outputs: the weighted average hue, the weighted average saturation, and the weighted average luminance.

6.2.3 Spatial Frequency's Radial Power

In the other processing pipeline, the image is converted to grey scale and its discrete Fourier transform is calculated, as introduced in section 2.3. The radial power of the spatial frequencies, which is simply the squared amplitude of the component frequencies, is then passed to the next stage. The spatial frequency pipeline also has three outputs:

the radial power of high spatial frequency, the radial power of median spatial frequency, and the radial power of low spatial frequency.

6.2.4 From Colour and Spatial Frequency to Affective Level

The colour pipeline and spatial frequency pipelines merge here. As introduced in section 2.5, HSL values and power of spatial frequencies can be mapped to affective levels using non-linear relations. Valtchanov used algorithms introduced in previous research (Fernandez & Wilkins, 2008; Juricevic et al., 2010; Torralba & Oliva, 2001; Valtchanov, 2013) to map HSL values from colour space to affective levels. Similarly, Valtchanov and Hancock used findings from previous research to map the radial power of spatial frequencies to affective levels as well (Fernandez & Wilkins, 2008; Juricevic et al., 2010; Torralba & Oliva, 2001; Valtchanov, 2013). These affective values are then fed into a neural network model in the next stage.

6.2.5 Customized Neural Network

This Unity plug-in provides a pre-trained neural network model that takes three inputs: the corresponding affective values of HSL values and power of spatial frequencies, the raw image data, and user profile (age, sex, nationality). The raw image data is processed by a pre-trained ResNet 50 model and the other two are fed into a multi-layer perception (MLP) network. This neural network returns two outputs: the expected arousal score and the expected pleasure score. Users can also replace the provided model with their own.

6.3 Summary

This chapter describes how to use our Unity plug-in to see the predicted arousal/pleasure score in the process and explains the algorithm used to build this plug-in. The next chapter concludes this thesis and discusses some limitations of our Unity plug-in.

Chapter 7

Conclusion & Limitations

In this thesis, we explored the possibility of developing tools for game designers to automatically calculate the expected affective valence (EAV) of game scenes using psychological characteristics of raw image data, including colour and spatial frequency. We improved Valtchanov and Hancock’s 2015 algorithm by building a linear regression model and a neural network model. We also compared the performance of using different methods and found that neural networks achieved the best prediction accuracy. The Unity plug-in we built can greatly help game designers better understand the emotional consequence of visual design in games.

7.1 Thesis Contributions

This thesis presents the following contributions:

1. We present empirical results from a survey study that collected individuals’ emotional reactions to game screenshots and explore possible variables that affect the satisfaction and arousal scores reported by participants. The data collected in our study is valuable for other researchers who are interested in people’s emotional responses to images specific to video games.
2. We use linear regression methods to construct a prediction model from the collected data. We validate the linearity assumption and measure the model’s accuracy.

3. We use supervised machine learning methods to construct a prediction model from the collected data. This machine learning can be integrated into our prototype and make the predictions offer more reliable accuracy.
4. We design and develop a Unity game engine plug-in, a prototype tool to help game developers check the real-time automated analysis of the expected affective valence of the game scene. This plug-in can act as a tool to assist designers, developers, researchers, and academics to embed affective constructs within new games and understand those employed in existing ones. We provide a formal approach to understanding the role emotions play in games at all levels and bridge the gap between game design, development and technology.

7.2 Limitations and Future Work

In our raw data collected in the online study, each image in the dataset was only rated once. As a consequence, the noise caused by random behaviour could not be ignored. As discussed in section 3.6, the data points on the box plot are quite scattered. Z -score normalization has also been applied to the dataset such that the mean of all of the values is 0 and the standard deviation is 1. However, there's no noticeable difference between before and after applying normalization. Noisy data may negatively impact neural network training. Therefore, future work would benefit from a more rigorous in-person study that collects participants emotional responses with less noise.

Furthermore, in the online study, participants only saw random screenshots from different video games. Therefore, they were not aware of the context of the game or the story told by the game. Story and narratives in the games are powerful tools to impact players' emotional experiences (Jiménez, 2016). However, the prediction models introduced in this thesis do not consider the effect of game stories. This Unity plug-in is limited to predicting emotional state based on colour and spatial frequency, and future work could consider incorporating contextual factors, such as the game's story, in a more refined neural network.

Our existing Unity plug-in can also be improved with new features. Currently, it only provides users with the predicted affective levels without any suggestions on specific approaches to achieve certain desired emotional states. Future efforts can be made to build a more complex neural network that can intelligently generate detailed instructions such as telling users what colours they should add and what objects they should place in game scenes.

References

- Amir, O., Biederman, I., & Hayworth, K. J. (2011). The neural basis for shape preferences. *Vision Research*, *51*(20).
- Bar, M., & Neta, M. (2006). Humans prefer curved visual objects. *Psychological Science*, *17*(8).
- Bar, M., & Neta, M. (2007). Visual elements of subjective preference modulate amygdala activation. *Neuropsychologia*, *45*(10).
- Barrett, L. F. (2014). The conceptual act theory: A précis. *Emotion Review*, *6*(4).
- Behnke, M., Chwilkowska, P., & Kaczmarek, L. D. (2021). What makes male gamers angry, sad, amused, and enthusiastic while playing violent video games? *Entertainment Computing*, *37*.
- Betella, A., & Verschure, P. F. (2016). The affective slider: A digital self-assessment scale for the measurement of human emotions. *PLoS ONE*, *11*(2).
- Bozinovski, S. (2020). Reminder of the first paper on transfer learning in neural networks, 1976. *Informatika (Slovenia)*, *44*(3).
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1).
- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Calleja, G., Herrewijn, L., & Poels, K. (2016). Affective Involvement in Digital Games.
- Carbon, C. C. (2010). The cycle of preference: Long-term dynamics of aesthetic appreciation. *Acta Psychologica*, *134*(2).
- Carlier, S., Van der Paelt, S., Ongenaes, F., De Backere, F., & De Turck, F. (2020). Empowering children with ASD and their parents: Design of a serious game for anxiety and stress reduction†. *Sensors (Switzerland)*, *20*(4).
- Cleveland, W. S. (1981). LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. *The American Statistician*, *35*(1).
- Dahouda, M. K., & Joe, I. (2021). A Deep-Learned Embedding Technique for Categorical Features Encoding. *IEEE Access*, *9*.

- Davis, J. P., Steury, K., & Pagulayan, R. (2005). A survey method for assessing perceptions of a game: The consumer playtest in game design. *Game Studies*, 5(1).
- de Byl, P. (2015). A conceptual affective design framework for the use of emotions in computer game design. *Cyberpsychology*, 9(3).
- Fernandez, D., & Wilkins, A. J. (2008). Uncomfortable images in art and nature. *Perception*, 37(7).
- Fintzi, A. R., & Mahon, B. Z. (2014). A bimodal tuning curve for spatial frequency across left and right human orbital frontal cortex during object recognition. *Cerebral Cortex*, 24(5).
- Gardner, E. S. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, 4(1).
- Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology*, 59.
- Hanjalic, A. (2006). Extracting Moods from Pictures and Sounds: Towards truly personalized TV. *IEEE Signal Processing Magazine*, 23(2).
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Springer Series in Statistics The Elements of Statistical Learning - Data Mining, Inference, and Prediction* (Vol. 2nd).
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December*.
- Jiménez, O. (2016). Affect During Instructional Video Game Learning: Story's Potential Role. *Emotions, Technology, and Digital Games*.
- Juricevic, I., Land, L., Wilkins, A., & Webster, M. A. (2010). Visual discomfort and natural image statistics. *Perception*, 39(7).
- Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Transactions on Neural Networks and Learning Systems*.
- Liljedahl, M. (2010). Sound for fantasy and freedom. *Game sound technology and player interaction: Concepts and developments*.
- Lindeman, R. H., Merenda, P. F., & Gold, R. Z. (1981). Introduction to Bivariate and Multivariate Analysis. *Journal of the American Statistical Association*, 76(375).
- Lucio de Mattos, R. (2020). Game Design for Emotions: a practicebased research on the design of game emotional experience. *Link Symposium Abstracts 2020*.
- Manjula, K., Spoorthi, S., Yashaswini, R., & Sharma, D. (2022). Plant Disease Detection Using Deep Learning. *Lecture Notes in Electrical Engineering*, 783.
- Naser, M. Z., & Alavi, A. H. (2021). Error Metrics and Performance Fitness Indicators for Artificial Intelligence and Machine Learning in Engineering and Sciences. *Architecture, Structures and Construction*.

- Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(19).
- Poels, K., Hoogen, W. V. D., Ijsselstein, W., & De Kort, Y. (2012). Pleasure to play, arousal to stay: The effect of player emotions on digital game preferences and playing time. *Cyberpsychology, Behavior, and Social Networking*, *15*(1).
- Porter, A. M., & Goolkasian, P. (2019). Video games and stress: How stress appraisals and game content affect cardiovascular and emotion outcomes. *Frontiers in Psychology*, *10*(APR).
- Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, *17*(3).
- Rosalind Picard. (1998). Affective Computing. *Trends in Cognitive Sciences*, *2*(7).
- Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, *11*(3).
- Stone, M. (1976). Cross-Validatory Choice and Assessment of Statistical Predictions (With Discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, *38*(1).
- Torralba, A., & Oliva, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3).
- Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*(3).
- Valdez, P., & Mehrabian, A. (1994). Effects of Color on Emotions. *Journal of Experimental Psychology: General*, *123*(4).
- Valtchanov, D. (2013). *Exploring the Restorative Effects of Nature: Testing A Proposed Visuospatial Theory* (tech. rep.).
- Valtchanov, D., & Hancock, M. (2015). Enviropulse: Providing feedback about the expected affective valence of the environment. *Conference on Human Factors in Computing Systems - Proceedings, 2015-April*, 2073–2082.
- Watson, D., Clark, L., & Tellegan, A. (1988). Worksheet 3.1 The Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) PANAS Questionnaire. *Journal of personality and social psychology*, *54*.
- Wilkins, A., Nimmo-smith, I., Tait, A., Mcmanus, C., Sala, S. D., Tilley, A., Arnold, K., Barrie, M., & Scott, S. (1984). A neurological basis for visual discomfort. *Brain*, *107*(4).
- Wilms, L., & Oberfeld, D. (2018). Color and emotion: effects of hue, saturation, and brightness. *Psychological Research*, *82*(5).
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*.

Zoph, B., & Le, Q. (2018). Searching for activation functions. *6th International Conference on Learning Representations, ICLR 2018 - Workshop Track Proceedings*, 1–13.

APPENDICES

Appendix A

R Studio Outputs

A.1 Summery and ANOVA of 12 Images' Arousal/Pleasure Scores

SUMMARY of arousal scores

Groups	Count	Sum	Average	Variance
Image 1	100	4856	48.56	845.764
Image 2	100	6159	61.59	787.7797
Image 3	100	5799	57.99	648.3938
Image 4	100	5284	53.37374	841.5834
Image 5	100	5097	50.97	734.0092
Image 6	100	5690	56.9	877.5253
Image 7	100	4164	41.64	641.9095
Image 8	100	5095	50.95	629.9268
Image 9	100	4824	48.24	698.8711
Image 10	100	3403	34.03	972.9789
Image 11	100	4377	43.77	1235.795
Image 12	100	5182	51.82	599.5834

(a) Summary of 12 Images' Arousal Scores

ANOVA of arousal scores

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	63111.35	11	5737.396	7.236856	4.81E-12	1.796703
Within Groups	941056.3	1187	792.8023			
Total	1004168	1198				

(b) ANOVA of 12 Images' Arousal Scores

Table A.1: Summary and ANOVA of 12 Images' Arousal Scores

SUMMARY of pleasure scores

Groups	Count	Sum	Average	Variance
Image 1	100	5547	55.47	762.8375
Image 2	100	6914	69.14	698.7681
Image 3	100	6987	69.87	564.8617
Image 4	100	5039	50.39	807.3716
Image 5	100	5242	52.42	751.4784
Image 6	100	5691	56.91	949.2544
Image 7	100	5039	50.39	758.907
Image 8	100	5490	54.9	685.6263
Image 9	100	4930	49.3	647
Image 10	100	3353	33.53	718.3324
Image 11	100	4821	48.21	1459.238
Image 12	100	4949	49.49	598.5555

(a) Summery of 12 Images' Pleasure Scores

ANOVA of pleasure scores

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	101074.5	11	9188.586	11.72733	4.67E-21	1.796696
Within Groups	930820.9	1188	783.5193			
Total	1031895	1199				

(b) ANOVA of 12 Images' Pleasure Scores

Table A.2: Summery and ANOVA of 12 Images' Pleasure Scores

A.2 Summery of linear regression models

```
1 lm(formula = Study_Arousal ~ Hue + Sat + Lum + HSF + MSF + LSF
2     ,
3     data = 1800_data)
4 Residuals:
5     Min       1Q   Median       3Q      Max
6 -54.86 -22.66   2.45   22.34   60.22
7
8 Coefficients:
9             Estimate Std. Error t value Pr(>|t|)
10 (Intercept) 30.341666   7.286607   4.164 3.28e-05 ***
11 Hue          0.003608   0.016969   0.213  0.8317
12 Sat          0.016234   0.024869   0.653  0.5140
13 Lum          0.032102   0.044673   0.719  0.4725
14 HSF         -0.214933   0.088761  -2.421  0.0156 *
15 MSF          0.157821   0.114306   1.381  0.1675
16 LSF          0.095347   0.091744   1.039  0.2988
17 ---
18 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
19
20 Residual standard error: 28.84 on 1793 degrees of freedom
21 Multiple R-squared:  0.006566, Adjusted R-squared:  0.003241
22 F-statistic: 1.975 on 6 and 1793 DF, p-value: 0.06601
```

Listing A.1: Summary of linear regression model to predict arousal score

```
1 lm(formula = Study_Pleasure ~ Hue + Sat + Lum + HSF + MSF +
2     LSF ,
3     data = 1800_data)
4 Residuals:
5     Min       1Q   Median       3Q      Max
6 -59.714 -22.564   3.269  23.036  58.777
7
8 Coefficients:
9             Estimate Std. Error t value Pr(>|t|)
```

```

10 (Intercept) 17.59472      7.37198      2.387 0.017103 *
11 Hue        -0.01126      0.01717     -0.656 0.512058
12 Sat         0.03784      0.02516      1.504 0.132816
13 Lum         0.12716      0.04520      2.813 0.004954 **
14 HSF        -0.33377      0.08980     -3.717 0.000208 ***
15 MSF         0.33286      0.11564      2.878 0.004046 **
16 LSF         0.08491      0.09282      0.915 0.360447
17 ---
18 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
19
20 Residual standard error: 29.18 on 1793 degrees of freedom
21 Multiple R-squared:  0.02309, Adjusted R-squared:  0.01982
22 F-statistic: 7.063 on 6 and 1793 DF,  p-value: 1.899e-07

```

Listing A.2: Summary of linear regression model to predict pleasure score

A.3 Neural network training process

A.3.1 MLP

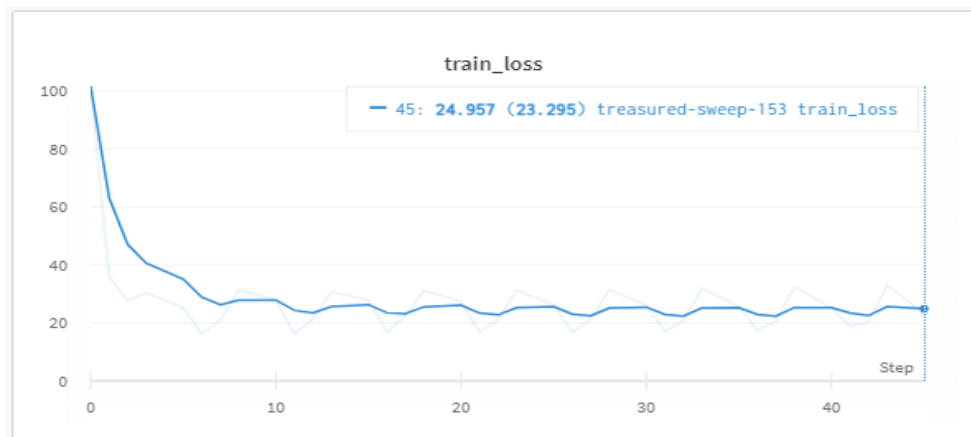


Figure A.1: Training process of MLP network for predicting pleasure score (smoothing value = 0.5).

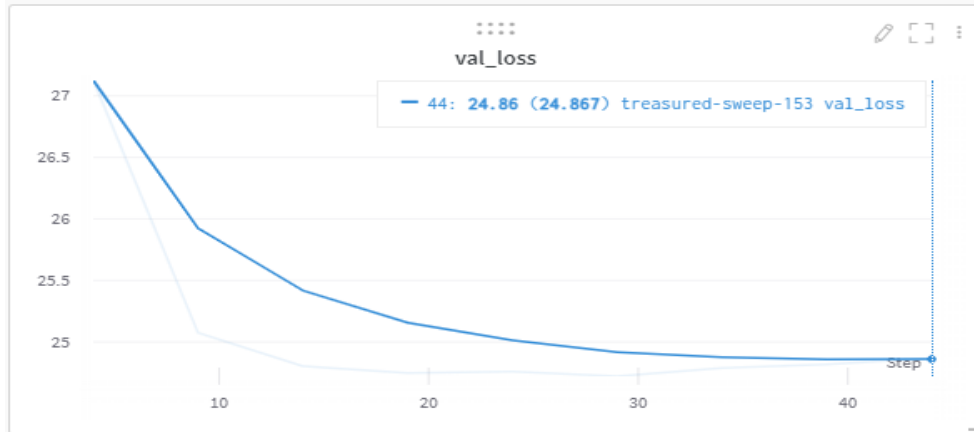


Figure A.2: Validation process of MLP network for predicting pleasure score (smoothing value = 0.5).



Figure A.3: Test process of MLP network for predicting pleasure score (smoothing value = 0.5).



Figure A.4: Training process of MLP network for predicting arousal score (smoothing value = 0.5).

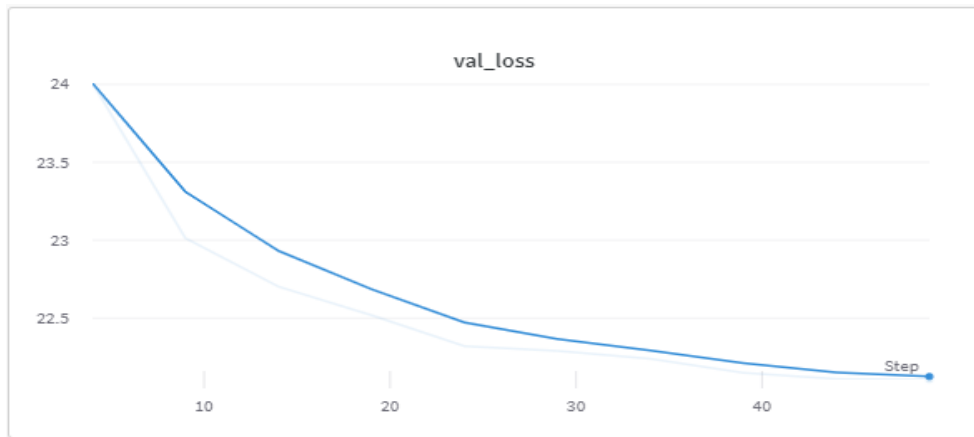


Figure A.5: Validation process of MLP network for predicting arousal score (smoothing value = 0.5).

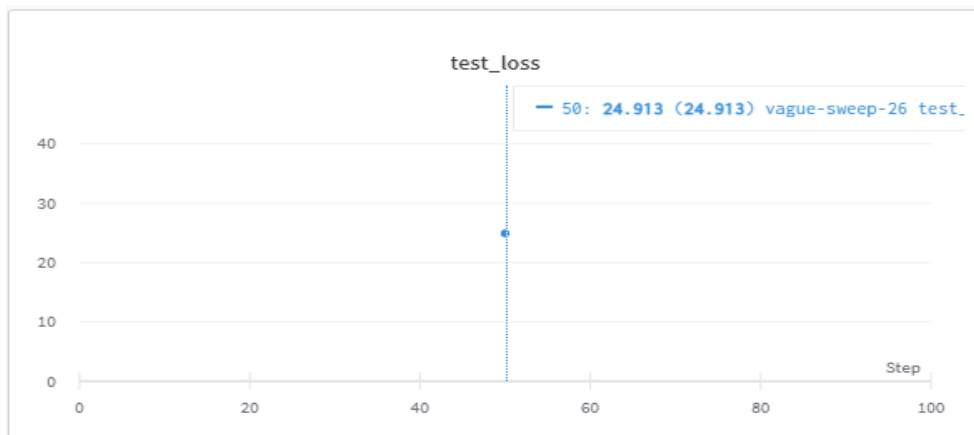


Figure A.6: Test process of MLP network for predicting arousal score (smoothing value = 0.5).

A.3.2 NLP + Resnet50

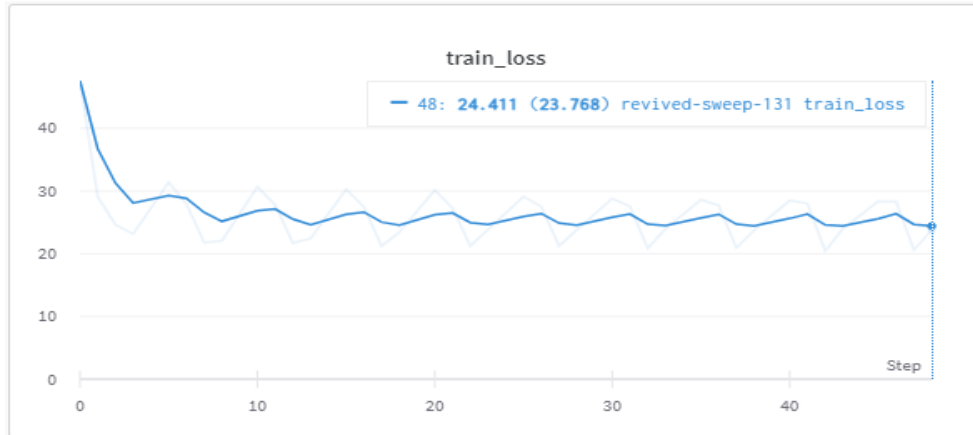


Figure A.7: Training process of MLP network combined with Resnet50 for predicting pleasure score (smoothing value = 0.5).



Figure A.8: Validation process of MLP network combined with Resnet50 for predicting pleasure score (smoothing value = 0.5).

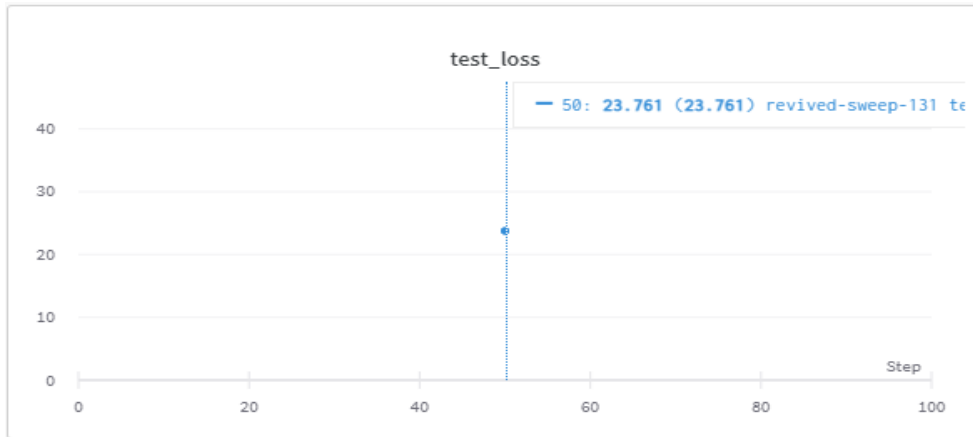


Figure A.9: Test process of MLP network combined with Resnet50 for predicting pleasure score (smoothing value = 0.5).

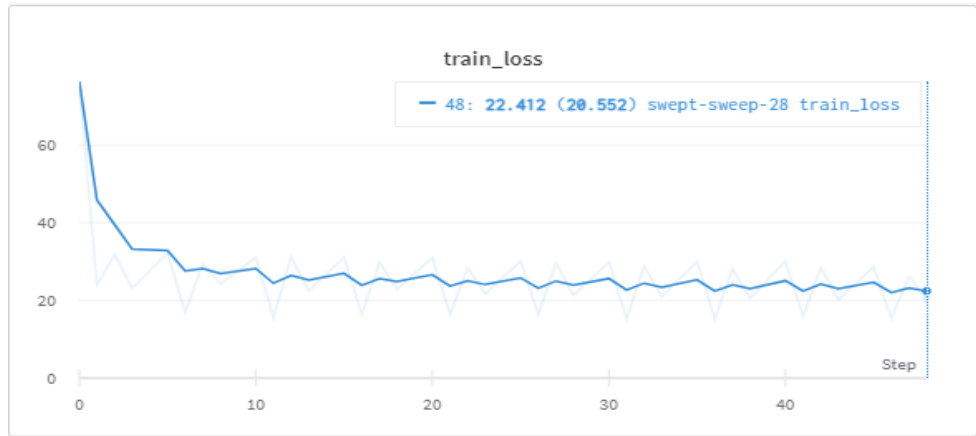


Figure A.10: Training process of MLP network combined with Resnet50 for predicting arousal score (smoothing value = 0.5).

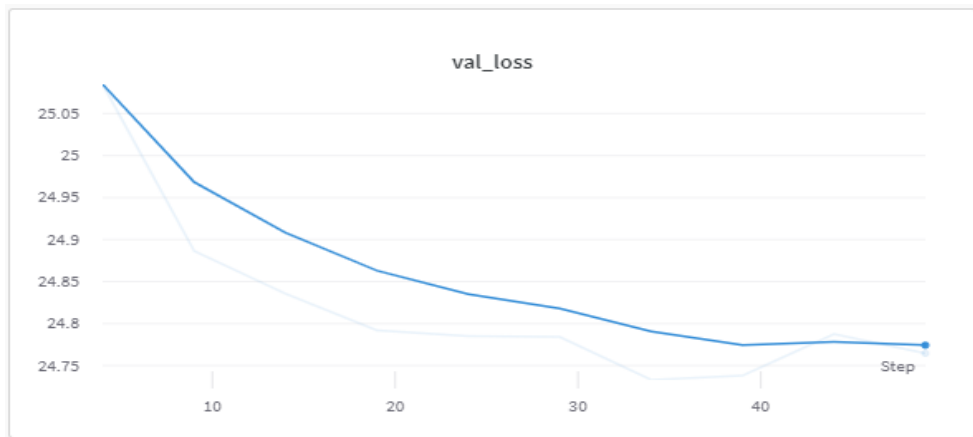


Figure A.11: Validation process of MLP network combined with Resnet50 for predicting arousal score (smoothing value = 0.5).

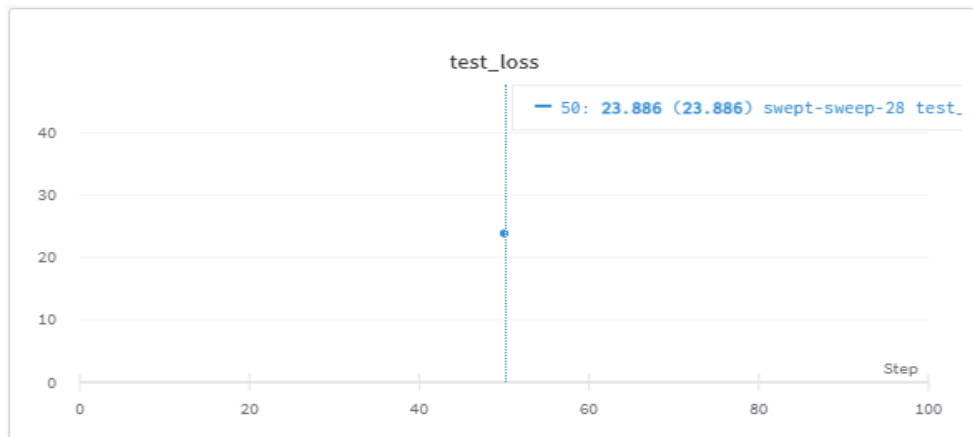


Figure A.12: Test process of MLP network combined with Resnet50 for predicting arousal score (smoothing value = 0.5).

A.3.3 NLP + Resnet50 + profile

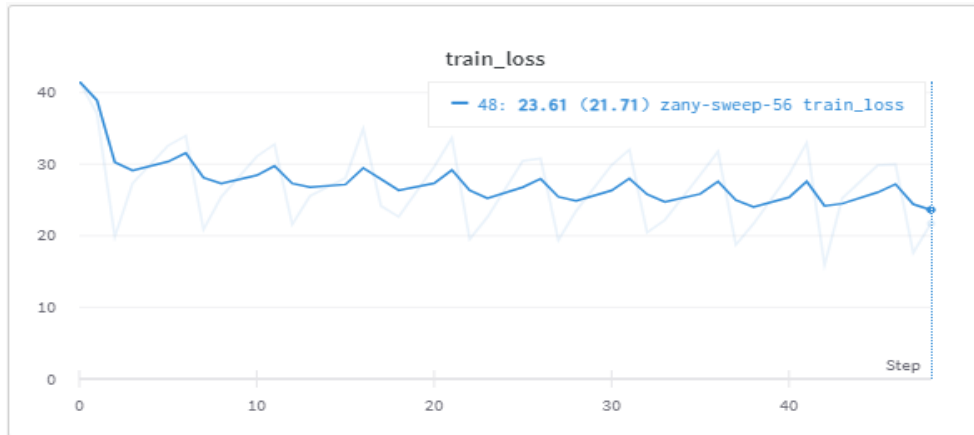


Figure A.13: Training process of MLP network combined with Resnet50 and user-profile integration for predicting pleasure score (smoothing value = 0.5).



Figure A.14: Validation process of MLP network combined with Resnet50 and user-profile integration for predicting pleasure score (smoothing value = 0.5).

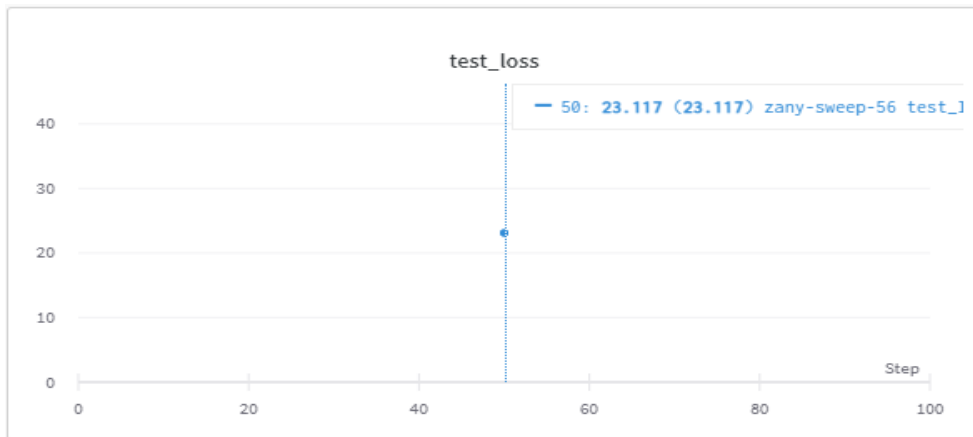


Figure A.15: Test process of MLP network combined with Resnet50 and user-profile integration for predicting pleasure score (smoothing value = 0.5).



Figure A.16: Training process of MLP network combined with Resnet50 and user-profile integration for predicting arousal score (smoothing value = 0.5).

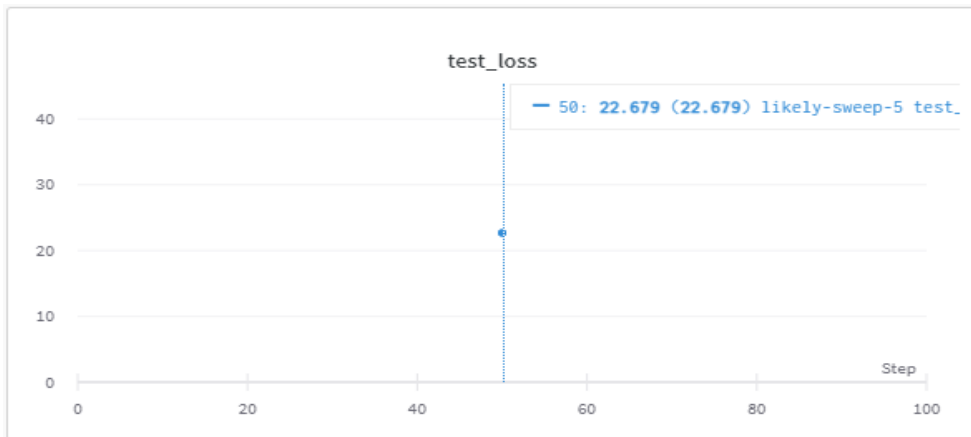


Figure A.17: Validation process of MLP network combined with Resnet50 and user-profile integration for predicting arousal score (smoothing value = 0.5).



Figure A.18: Test process of MLP network combined with Resnet50 and user-profile integration for predicting arousal score (smoothing value = 0.5).