УНИВЕРЗИТЕТ У НОВОМ САДУ

ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА У НОВОМ САДУ

Цвета Мајтановић

Аутоматско повећање памтљивости слика

ДОКТОРСКА ДИСЕРТАЦИЈА

Нови Сад, 2021.

Automatic Enhancement of Image Memorability

DOCTORAL DISSERTATION

Cveta Majtanović

Faculty of Technical Sciences

University of Novi Sad

Novi Sad, 2021

УНИВЕРЗИТЕТ У НОВОМ САДУ ● **ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА**
21000 НОВИ САД, Трг Доситеја Обрадовића 6

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

Редни број, **РБР**:	
Идентификациони број, **ИБР**:	
Тип документације, **ТД**:	Монографска документација
Тип записа, **ТЗ**:	Текстуални штампани материјал
Врста рада, **ВР**:	Докторска дисертација
Аутор, **АУ**:	Цвета Мајтановић
Ментор, **МН**:	Др Дубравко Ћулибрк, др Нику Себе
Наслов рада, **НР**:	АУТОМАТСКО ПОВЕЋАЊЕ ПАМТЉИВОСТИ СЛИКА
Језик публикације, **ЈП**:	Енглески
Језик извода, **ЈИ**:	Српски / Енглески
Земља публиковања, **ЗП**:	Република Србија
Уже географско подручје, **УГП**:	АП Војводина, Нови Сад
Година, **ГО**:	2021.
Издавач, **ИЗ**:	Ауторски репринт
Место и адреса, **МА**:	Факултет техничких наука, 21000 Н. Сад, Трг Доситеја Обрадовића 6
Физички опис рада, **ФО**: (поглавља/страна/ цитата/табела/слика/графика/прилога)	5 поглавља / 113 страна / 102 цитата / 31 слике / 7 табела / 1 прилог
Научна област, **НО**:	Индустријско инжењерство и инжењерски менаџмент
Научна дисциплина, **НД**:	Инжерњерски менаџмент
Предметна одредница/Кqучне речи, **ПО**:	Машинско учење, Рачунарска визија, Вештачке неуронске мреже
УДК	Монографска документација
Чува се, **ЧУ**:	Библиотека Факултета техничких наука у Новом Саду
Важна напомена, **ВН**:	
Извод, **ИЗ**:	Дисертација разматра проблем аутоматског повећања памтљивости фотографије на основу модела дубоког учења. Овој проблематици се приступа са аспекта развоја иновативног приступа заснованог на парадигми уређивања слике применом филтера. Арбитрарна улазна слика аутоматски преузима сет стилских карактеристика који се преносе путем алгоритма неуронског стила, омогућавајући на овај начин пораст памтљивости целокупне слике. Ефикасност предложеног приступа евалуирана је експериментално уз изведбу корисничке студије.
Датум прихватања теме, **ДП**:	31.10.2019.
Датум одбране, **ДО**:	

Чланови комисије, **КО**:	Председник:	Проф. др Владимир Божовић, редовни професор
	Члан:	Проф. др Никулае Себе, редовни професор
	Члан:	Проф. др Борко Фурхт, редовни професор
	Члан:	Др Мирсад Ћосовић, доцент
	Члан, ментор:	Проф. др Дубравко Ћулибрк, редовни професор

Потпис ментора

УНИВЕРЗИТЕТ У НОВОМ САДУ ● **ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА**
21000 НОВИ САД, Трг Доситеја Обрадовића 6

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

Accession number, **ANO**:	
Identification number, **INO**:	
Document type, **DT**:	Monograph type
Type of record, **TR**:	Printed text
Contents code, **CC**:	PhD dissertation
Author, **AU**:	Cveta Majtanovic
Mentor, **MN**:	Dubravko Culibrk, PhD, Full professor
Title, **TI**:	Automatic Enhancement of Image Memorability
Language of text, **LT**:	English
Language of abstract, **LA**:	Serbian / English
Country of publication, **CP**:	Republic of Serbia
Locality of publication, **LP**:	Autonomous Province of Vojvodina
Publication year, **PY**:	2021
Publisher, **PB**:	Author's reprint
Publication place, **PP**:	Faculty of Technical Science, Novi Sad
Physical description, **PD**: (chapters/pages/ref./tables/pictures/graphs/appendixes)	5 chapters / 113 pages / 102 citations / 31 images / 7 tables / 1 appendix
Scientific field, **SF**:	Industrial Engineering and Engineering Management
Scientific discipline, **SD**:	Engineering Management
Subject/Key words, **S/KW**:	Machine Learning, Computer Vision, Artificial Neural Network
UC	
Holding data, **HD**:	The Library of the Faculty of Technical Science, Novi Sad
Note, **N**:	
Abstract, **AB**:	The dissertation considers the problem of automatic increase of image memorability. The problem-solving approach is based on editing-by-applying-filters paradigm. Given an arbitrary input image, the proposed deep learning model is able to automatically retrieve a set of "style seeds", i.e., a set of style images which, applied to the input image through a neural style transfer algorithm, provide the highest increase in memorability. We show the effectiveness of the approach with experiments, performing both a quantitative evaluation and a user study.
Accepted by the Scientific Board on, **ASB**:	31.10.2019
Defended on, **DE**:	

Defended Board, **DB**:	President:	Vladimir Božović, PhD, Full professor	
	Member:	Niculae Sebe, PhD, Full professor	
	Member:	Borko Furht, PhD, Full professor	
	Member:	Mirsad Ćosović, PhD, Docent	Menthor's sign
	Member, Mentor:	Ćulibrk Dubravko, PhD, Full professor	

Acknowledgments

Undertaking this PhD has truly been a life-changing experience and it would not have been possible without the support and guidance that I received for many incredible people.

I gratefully acknowledge my supervisor prof. dr Dubravko Ćulibrk for his incredible optimism, for being visionary and welcoming me warmly into the Computer Science community, for all the encouragements, continuous support and advices he gave me over the last couple of years, for being able to transfer his calmness and positive life attitude into every activity of our working environment. I am grateful for his exceptional acts of kindness and generosity, motivation, knowledge and expertise, without his existence, guidance and constant feedbacks, this PhD would not have been achievable. I could not have imagined having a better advisor and mentor.

Many sincere thanks to my other supervisor, prof. dr Nicu Sebe for gathering the most inspiring people around goals addressed in a large spectrum of themes within the MHUG (Multimedia and Human Understanding Group) of the Department of Information Engineering and Computer Science at the University of Trento. Among these people, special and big thanks to my dear colleague and friend Gloria Zen, my Python and "zen" teacher who helped me and inspired in numerous ways during various stages of my PhD.

Thanks to my two incredible and beloved Universities of Novi Sad and Trento, for giving me the access to the laboratories and research facilities and for making this interdisciplinary work possible. To all professors and staff.

I would also like to thank my lovely colleagues and friends, above all my dearest Bojana Bokan for believing in me and supporting my first steps. To all my colleagues and lab-mates in Povo 2, SanBa and back in Mashinac for the most stimulating discussions, to Bojana Milić for everything she is. To Stefano, for the most incredible support, inspiration and love.

Finally, I would like to thank my parents for setting the course for all this long time ago.

Cveta Majtanović,
In Trento/Novi Sad, 2021

Abstract

Every picture tells a story.

Images are one of the most dominant types of media, uploaded an average in billions every single day and in hundreds of billions on an annual basis. Artifacts depicting visual perception like photographs and other two-dimensional pictures are distributed through the growing number of image-sharing websites. Consequently, a thriving interest in understanding the whole image or objects depicted in it, its style or the emotions a picture might evoke, together with all the other image properties, became increasingly represented in research practice. This research focuses on the problem of automatically enhancing memorability of an image.

Recent works in Computer Vision and Multimedia have shown that intrinsic image properties like memorability can be automatically inferred by exploiting powerful deep learning models. This research advances the state of the art in this area by addressing a novel and more challenging issue: "*Can we transform an arbitrary input image and make it more memorable?*". To state this question properly one requires the existence of memorability measures. Methods for automatically increasing image memorability would have an impact in many application fields, such as education, gaming or advertising.

To tackle the problem, we introduce an approach inspired by editing-by-applying-filters paradigm, adopted in photo editing applications like Instagram and Prisma. Users of the two apps generally have to go through the available filters before finding the desired solution which is turning the editing process into a resource- and time-consuming task. In the work conducted for the purpose of this thesis, we reverse the process: given an input image, we propose to automatically retrieve a set of "style seeds", i.e., a set of style images which, applied to the input image through a neural style transfer algorithm, provide the highest increase in memorability. As a result, we demonstrate that it is possible to automatically retrieve the best style seeds for a given image, thus, remarkably reducing the number of human attempts needed to find a good match.

Furthermore, we show the effectiveness of the proposed approach with experiments on the publicly available LaMem dataset, performing both a quantitative evaluation and a user study. To demonstrate the flexibility of the proposed framework, we also analyze the impact of different implementation choices, such as using different state of the art neural style transfer methods. Finally, we show several qualitative results to provide additional insights on the link between image style and memorability.

This approach arises from recent advances in the field of image synthesis and adopts a deep architecture for generating a memorable picture from a given input image and a style seed. Importantly, to automatically select the best style, also relying on deep models, a novel learning-based solution is proposed. The experimental evaluation, conducted on publicly available benchmarks, demonstrates the effectiveness of the proposed approach for generating memorable images through automatic style seed selection.

Contents

Chapter 1

INTRODUCTION

The explosive growth of multimedia market, with a simultaneous increase in the amount of multimedia content, particularly photos and videos generated and shared by users in different web-based collections, has inspired a change in user-centered design (UCD) approach. As an iterative design process of building insights about users' experience, the UCD requires constant examination about the ways in which users are likely to consume a certain product (i.e., to use an application) and consequently, it inspires the development of new approaches in designing algorithms. For illustrative purposes, by the end of 2021 video will account for 80% of internet traffic [1] and it has been reported that the number of video content hours uploaded every 60 seconds grew by around 40% between 2014 and 2019[1]. Also, as of today more than 50 billion photos and videos have been uploaded on the popular Instagram photo sharing platform.

In addition to the number of positive effects, rapid flow of data has inevitably led to an overload of content that people are exposed to, on daily basis. In these circumstances, the possibility of retaining, reproducing, in other words remembering all that content and the information presented this way is becoming more of a problem the growing scientific importance attached to research and exploration of the field.

An idiom according to which "*A picture is worth a thousand words*" (i.e., 84.1 words [2]), the allusion to the fact that a complex idea can be explained with only one image, has probably inspired the emergence of information graphics, known as infographics. With this form of visual representation of information, data and knowledge, graphic designers and marketers have created the ability to convey complex information faster and in a clearer way. Nevertheless, many research questions are still open. Finding modalities for attracting and retaining subject's attention, while increasing the likelihood of re-recognizing once-seen content (memorability) could have number of practical applications in many disciplines.

This dissertation examines ways to automatically increase the efficiency of absorbing information presented in a visual form (i.e., photography) by manipulating the style of a single photo. It explores also possibilities of using an increasing amount of publicly available image

[1] https://www.statista.com/

datasets, while finding new machine learning models to increase an overall image memorability, i.e., the chance of the content of a given photograph to be better remembered.

This thesis arose at the intersection of several fields, involving a number of novel findings about image memorability coming both from psychological sciences and organized human experiments and from developed computational models. Knowing that every single person has a unique set of individual experiences forming their memories, it can be expected that these individual differences affect the final decision when memorizing content. Surprisingly, recent studies have shown that people tend to both remember and forget the same things and this happens because images differ in their memorability. These variations are caused by intrinsic features of an image which reveal how memorable a particular image will be across different observers. Image Memorability is, therefore, an objective quantitative measure of an image, it is independent of the observer and can be computationally predicted.

This consistency across viewers opened-up a possibility for numerous computational applications whose results could impact fields such as education, when selecting study material, marketing and design of promotional material, visual design and memorability of informational graphics, or Human-Computer Interaction (HCI) where could progress the usability goals, the end-user experience and satisfaction.

1.1. Motivation

Life in an instant society where each activity's imperatives are “quick and fast” and where the amount of visual stimuli humans are exposed to keeps growing, tackled the question about an average adult human brain's information storage capacity. Not only that it tackles the challenge of quantifying the range of gigabytes for data storage, but it also questions the selection criteria of a human brain when it comes to retaining knowledge derived from different sources. It has been noted that the brain has a tendency to absorb and maintain the knowledge derived from images and videos, favoring it over the other types of signals [3]. This discovery found an immediate implementation in marketing, when graphic designers started visually representing both textual information and data in upgraded chart forms and diagrams known as infographics.

However, this superiority effect is not applicable to the comparison of two photographs. Cases in which some images are easy to remember, while others are easily forgotten [4]. In consequence, marketers and graphic designers began a new struggle to capture consumers' visual attention in order to leave a trace in the recipient's memory. All this progressed the development of several digital tools such as Instagram and Prisma as a response to the "information-overload age" challenges. Designed to facilitate the modification of visual contents for capturing user gaze, their end-goal is certainly improving an overall image 'shareability', 'likability' and their memorability.

To overcome and tackle the challenge, marketers and designers are constantly searching state-of-the-art research, extracting general rules for reshaping visual contents and making them visually more appealing and easier to remember. It all starts from acquiring the knowledge about the way our brain responds to certain stimuli, insights delivered by marketing and branding theories.

It would be desirable to reduce human intervention and discover innovative technological solutions to automatically modify images and videos according to specific properties or attributes. In case of automatic enhancement of memorability, not only it would have an impact on direct commercial employment in marketing, but also, considering the way visual content influences people, a discovery of this kind would find its applications in education and knowledge dissemination.

1.2. Applications and Challenges

As images (and videos) became ubiquitous in production, distribution and consumption, their utilization and manipulation became necessary and even largely significant in different industries, like fashion industry, e-commerce websites, real estate agencies, online education, medicine, as well as in various social media and marketing activities. Moreover, the Global Photo Editing Software Market size ($843.8 million in 2019) is expected to reach 1.485.2 million by the end of 2027[2]. Also, image and video detection and manipulation programs became an integral part of governmental projects in many countries. There are different reasons for the initial motivation to change an image content or its style, either to make it more attractive according to some criteria, or to give it a completely different view from the original one, or, as in case of image editing apps, to increase the end-user engagement, largely depending on the purpose and application.

This work introduces an approach to automatically modify images to increase their chance to be remembered by the observers. Numerous research fields are addressing aspects of this problem, each one from its own perspective:

- *Cognitive Science*: behavioral aspects of brain's reaction to stimuli, like surprise learning [5], have been widely explored as well as the effects of successive and simultaneous information presentation on learner's visual search ability and working memory load for different information densities [6], since the processing of information in the brain depends on the capacity of visual short-term memory (VSTM).

- *Neuroscience*: focuses on the brain mapping when stimuli are present and when parts of the brain are involved in specific activities, claiming that this unlocks amygdala's mechanism for emotional learning and memory [7], such as in unpleasant situations when facing one's fears. Other works are focused on understanding which neural structures are emphasizing the visual memory capacity and the ways to preserve it.

- *Psychology*: several experimental studies [8] examined the capacity of recognition memory for both pictures and words, based on forced-choice recognition and measuring the retrieval speed, as well as the concept of memorability and its relation to other aspects of the human mind studied in [9].

- *Computer Science*: previous research studies on Artificial Intelligence and Computer Vision have shown that memorability is an intrinsic image property [10], implicating that different viewers demonstrate similar performance in remembering or forgetting

[2] https://www.businesswire.com/

specific images [11]. Other works demonstrated that image memorability can be automatically predicted [12] by employing advanced deep learning models with accuracy close to human performance.

All these works have applied different methodologies in order to address a number of common challenges and emphasize the importance of the problems they explored such as:

- Investigating the relationship between emotions and memorability [13], as well as the way it regulates consciousness, attention, and information processing.

- Addressing the limitations of the human visual system in computational cost terms.

- Discovering brain's ability to process images more quickly compared to verbal or written information.

1.3. Thesis Contribution and Structure

A successful visual advertising campaign implies its specificity and noticeability, where memorability plays a significant role. Graphic design, as the process of visual communication and problem-solving activity through the use of typography, photography and illustrations, recognizes the importance of these properties when it comes to creating visual identifiers, so-called logos. As a rule, a logo should convey the information about the type of business or product, and consequently provide a broader picture of corporate visual identity. This graphic element is the essence of brand identity and affected by the psychology of font, color and shape - details which will influence people's purchasing decisions - which is why much attention is regularly paid to its design. Yet, the time and creative resources invested in designing a visual image that will be noticeable and memorable are often ineffective, while the outcome and the degree of memorability of the image remains uncertain. A similar thing happens during the process of creating advertisements and promotional material, where images are the most dominant representations.

Not surprisingly, due to the great responsibility marketing managers and graphic designers have in the process of creating an image, which will be used to represent an advertising campaign, their motivation for storytelling improvements is quite extensive. Marketers are in charge of handling both financial and human resources in a project and they are also responsible for the end-result's level of visibility and memorability. Consequently, they are constantly searching for the most innovative ways to retain users' visual attention and leave an impression on the observers' memory, along with exploring new modalities for customer engagement. Because of this, the ability to measure and more importantly to increase memorability of an already-created image, could have a wide range of practical applications.

All this has opened up a number of research questions on the possibilities of image modification increase their memorability. Despite the fact that a considerable number of studies dealing with similar research issues has been conducted, no model has yet been found to automatically increase the memorability of visual content. This inspired us to try to design this research and attempt to solve the challenge.

We translate our ambitions in this automatic-memorability-increase model creation into the following problem statement: *is it possible to increase image memorability of an arbitrarily selected image while preserving its high-level content*?

A practical example which could help clarify this research question is the following: an advertising campaign concerning the design of a new product targeting a specific market sector. Once the very expensive design phase is over, the client receives a set of images which are promoting and advertising the new product. However, this client might be unsatisfied with the proposed images, while marketers have already invested all the resources and spent the budget on creating those images. So, they need an additional sales argument, i.e., *this image is more likely to be remembered*. Such images tell a story, meaning that in the attempt of increasing an image's memorability, the high-level *content*, that is the meaning, should remain intact.

In this work we propose a novel approach for increasing the memorability of images which is inspired by the editing-by-filtering framework. Built-in editing tools are largely adopted in communities similar to Instagram where the user engagement widely depends on applying a particular set of stylistic aesthetic choices. Once the user takes or uploads a photo (or video), he or she can edit this image and apply one of the existing filters by tapping the one they wish to use and then saving all the changes they made. However, users here have to investigate a while and try out all possible options before they identify that one filter they choose to use, which consumes a lot of time. Finally, even after they generate a new picture, under no circumstances can they know for sure whether it will be memorable, which inspired us to explore the reverse process: instead of going through the list of filters, could we recommend a set of style images that the user can apply to the input image in order to have as a result a slightly adjusted version with higher chances to be remembered. This procedure could notably shorten the editing time and cut down the number of human attempts necessary for the complete action.

To clarify the design steps of our approach, we need to outline (Fig. 1.1) the concept of Deep Learning, a subset of Machine Learning and Artificial Intelligence.

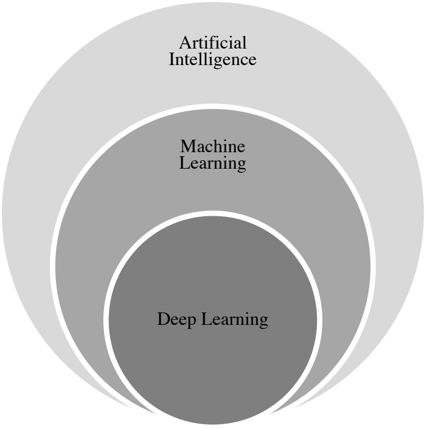


Figure 1.1: *The relationship between Artificial Intelligence, Machine Learning and Deep Learning*

Artificial Intelligence is a technique that enables machines to mimic human behavior. It represents the development of computer systems able to perform tasks which normally require human intelligence [14] i.e., visual perception, speech recognition, decision making, etc.

Machine Learning is a technique to achieve Artificial Intelligence (AI) through algorithms trained with data. Computers use huge amounts of data to learn how to perform certain tasks and build a model based on sample data (training data) with the aim of making predictions and taking decisions [15].

Deep Learning (known also as Deep Structured Learning) is a subset of Machine Learning inspired by the structure of the human brain which is here called an Artificial Neural Network. It has ability to learn without human supervision, drawing from structured and unstructured data. More precisely, learning can be supervised, semi-supervised or completely unsupervised. There are various Deep Learning architectures [15] like Deep Neural Networks, Recurrent Neural Networks, Convolutional Neural Networks and others used and applied to fields like Computer Vision, Speech Recognition, Natural Language Processing, Audio Recognition, etc.

Our methodology relies on Deep Neural Networks (DNN), Artificial Neural Networks with multiple layers between the input and output layers, trained to recognize various features, calculate probabilities and predict different characteristics or qualities, like users' personality from Instagram pictures [16] or the impact of the aesthetic value of an image to its popularity [17]. Our approach is extensively evaluated on the publicly available Large-Scale Image Memorability (LaMem) dataset, the largest annotated image memorability dataset to date.

This thesis makes contributions to both human vision and computer vision, by adding the understanding of human attention about images and by introducing an approach for increasing memorability of images. The main contributions of our work are the following:

- We introduce a novel framework to increase the memorability of images by tackling the task of increasing image memorability while keeping the high-level content intact, thus, modifying only the style of the image.

- We cast this into a style-based image synthesis problem of selecting the most memoralizable style "seeds" for a given image. In this way we keep the users in the loop allowing them to choose among a small set of top styles. The effectiveness of our approach in recommending the best styles is demonstrated with several qualitative and quantitative results and with a user study.

- We propose an automatic method to retrieve a set of style images, which are expected to lead to the largest increase of memorability for the input image and we demonstrate the flexibility of our framework with respect to different implementation choices.

- We propose a lightweight solution for training the network which will be used to retrieve the "best" set of style images, selecting them to efficiently learn our model with a reduced number of training data. Moreover, our approach is able to compute not only the best styles but also to determine the optimal degree of stylization.

The remainder of this thesis is structured as follows:

- Chapter 2 provides a review of research literature relevant to the work presented in this thesis.

- Chapter 3 presents the methodology used while conducting research for the purpose of this thesis, with an overview of tools developed to automatically increase the memorability of photographs by changing and modifying their style.

- Chapter 4 provides an overview of experiments conducted in various different scenarios using different sets of data, while determining the best model for predicting memorability by defining an experimental dataset, training and steps for testing the model. The efficiency of the proposed research is evaluated through experiments on a publicly accessible database.

- Chapter 5 is dedicated to the discussion, implications and concluding considerations.

Chapter 2

BACKGROUND AND RELATED WORK

Recognized as a significant characteristic of visual content, memorability has a strong potential and capacity to make photographs of certain products (or the brand itself) more effective. Discovering and exploiting AI-based solutions to craft compelling visual messages that communicate and deliver information in a more memorable way could have a significant impact on any advertising campaign. This thesis aims to contribute to the research focused on understanding the process humans use to remember visual data. The importance of visualization from a human perspective is significant, especially if we know that almost half of the brain is dedicated to processing visual information [18], directly or indirectly and that at least 65% of people are visual learners, according to the Social Science Research Network [19].

We will also review the existing memorability measures used to evaluate image memorability objectively, explore tools that have already been developed with the goal of altering the memorability of visual data and conduct experimental research defining more precisely the relationship between memorability and other properties, so that we can finally provide not only automatic prediction of memorability, but also its overall increase.

2.1. Overview and Objectives

The concept of memorability and its relations with other aspects of the human mind has been studied from a psychological [20, 21, 22], cognitive [9] and neuroscience [7, 23] perspectives. Computer science, on the other hand, mostly contributed with the explorations of image properties and their relation to memorability, along with some other subjective properties of visual data, such as interestingness, emotions evoked, aesthetics, colors [24] etc.

Although there are many areas exploring memorability in general, from various perspectives, there are three main research lines related to the work presented here:

- *Studies addressing automatic image manipulation for altering perceptual attributes.*
- *Studies analyzing visual memorability.*
- *Works on neural style transfer.*

Since each of these categories has been approached from different scientific standpoints, we will provide an overview of the most recent papers, organized by category.

2.2. Memorability and Related Perceptual Attributes

Every scientific discipline evaluates and investigates memorability and its relationship with related perceptual attributes from a different perspective.

2.2.1. Psychological Perspective and Investigations in Cognitive Science and Neuroscience

Visual memory has been investigated by several works, both from psychological [25, 26, 27] and computational perspective [28, 29]. Psychology focused mostly on the capacity of human memory in the context of encoding entities and how much attentional resources we dedicate to one specific task. Researchers discovered that the likelihood of encoding a given entity, or in other words – memorability trend, is highly consistent across viewers and intrinsic to an image. This means that people commonly have the tendency to remember and forget the same things when they come across someone or something new (seeing it for the first time). What also influences our memories goes beyond the memorability of the stimulus itself, like the intensity to grab our attention the stimulus has, but also some rather subjective elements, such as personal motivation and resources we dedicate to one specific task.

Psychology and cognitive science also explored how memorability interacts with various phenomena by running psychophysical and neurophysiological experiments to explore the link between memorability and other attention-related properties. In [30] authors emphasized the common neural architecture of the bottom-up (reflexive and typically driven by the properties of the observed objects) and top-down (voluntary and goal-driven) orientations to attention. They demonstrated that memorability remains resilient to all aspects and sensations [31]. Thus, it is truly an intrinsic attribute of an image.

Computer vision research also provided support for treating memorability as an intrinsic property of the visual content. Several works of Isola et al. examined intrinsic and extrinsic effects on image memorability [11]. They discovered that properties like distinctiveness [11, 8] and arousal [13, 7, 23] have a positive influence on memorability.

Inspired by cognitive and psychological research, works and studies of attention and visual memory are found to a large extent in computer vision, for instance, studies of human capacity for remembering object details, described in [20] or the effect of emotions on memory in [9]. Some of them studied the brain's learning mechanisms as well. E.g., the role of amygdala in memory. One of the most important discoveries is that human memory is imperfect, imprecise and subject to interference; despite the fact that observers are able to remember thousands of images, these memories lack detail. Conversely, researchers in [20] showed that long-term memory is capable of storing a massive number of objects with details from the image – participants of the study were correctly reporting the old item and achieving a remarkably high performance on 92% of the trials, which demonstrates that participants successfully maintained detailed representations of thousands of images. The significance of these results lies in their implications for cognitive models in which capacity limitations impose a primary computational constraint (e.g., models of object recognition) and a challenge to neural models of memory storage and retrieval.

Reflecting on number of scientific and research achievements, primarily in the field of cognitive science, neuroscience and psychology, a growing interest in examining the various properties of visual media has been noted. Among the general properties that affect the way a photo is perceived, such as quality, color and transparency, there is also the degree of picture's memorability. Since our research starts from the paradigm of editing photos with filters, we were also interested to know the way psychological scientific community examined this and we found a study in [32] which tested also the psychological reaction to the edited changes within a photo using post-processing techniques, while the Japanese Psychological Research published in [33] focused only on exploring the effectiveness of color in picture recognition memory. After manipulating color saturation of the image, or in other words – presenting the image with its original colors and also in its black and white version, it was discovered that the recognition performance was higher with colors. These were only the beginnings of the image manipulation tests from both technical and psychological sides, with discoveries that only stimulated further research in this direction. Today, almost 20 years later, we are witnessing the implementation of such findings in various fields of industry using marketing tools and surprise factors to arouse attention and stimulate consumers' impulses and, consequently, to activate their willingness to buy a particular product shown in a given photograph.

In recent years, the focus of researchers has been primarily on perception [34, 26], attention [3], and emotional experiences when exposed to some visual content [13]. Berlyne [35] has stated that there are two kinds of exploratory behavior, namely specific and diversive curiosity and has cited data relating to pleasingness and interestingness of complex visual patterns in support of his position. Eisenman [34] was testing this by varying degrees of complexity in terms of both -pleasingness and interestingness- and has confirmed that highly complex polygons are perceived as more interesting and the less complex polygons more pleasing.

Furthermore, Standing et al. [26] investigated perception and memory for pictures in an experimental study by using a two-alternative forced-choice task and discovered that the image presentation time could be reduced to one second per picture without seriously affecting performance, as well as that the stimuli could be reversed in orientation in the test situation without damaging recognition performance.

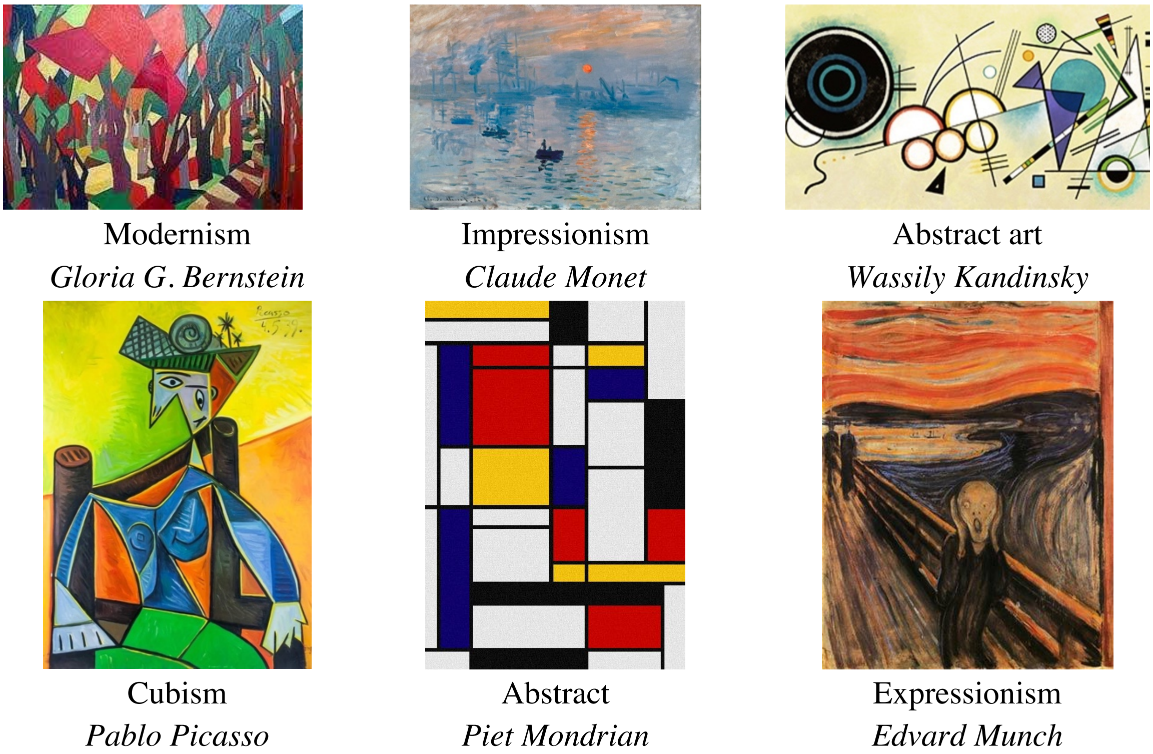
The relationship between emotional arousal and long-term memory has been investigated by exposing observers to emotionally relatively neutral short stories presented in slideshow and an emotionally arousing story. These tests were repeated after a period of time and also varied in different experimental situations and have shown that subjects who viewed the arousal story both experienced a greater emotional reaction to the story than did the subjects who viewed the neutral story, and subsequently exhibited enhanced memory for the story. Moreover, they also managed to recall more slides than the subjects who saw the neutral one, which implies that emotional arousal influences long-term memory in humans.

The number of publications emerging from the intersection of cognitive studies, psychology, computer science and information technology continues to grow, finding its applications also in the field of computer science. The study of Isola, et al. [10] provided strong evidence that memorability is an intrinsic property of the image, as opposed to the opinion prevailing in science until that moment. This means that different observers of the same photograph show approximately the same performance in the process of remembering and forgetting, which challenged previously valid views in science, which underlined the importance of individual differences [36]. This finding had a powerful impact on the cognitive theories, which used to stress the importance of individual differences, explaining how individuals may pay different levels of attention and employ different strategies and schemas for encoding memories differently [37, 38], which also had a crucial role in the different topic-related theories. These studies relied on memory pair games to provide an objective evaluation of image memorability, which had a surprisingly low variance across trials.

Other works showed a negative correlation between memorability and other image properties, such as its aesthetic value [39, 40]and interestingness [41]. More precisely, Aitken et al. found that visual interestingness positively correlates with the perceived image naturalness [42, 29] and complexity [43, 35, 34, 44]. These studies have provided tools to detect the visual features responsible for both memorable and easily-forgettable images. For instance, images that tend to be forgotten lack distinctiveness, like natural landscapes, whereas pictures with people, specific actions and events, or central objects, are reported to be far more memorable.

Works in psychology [28] investigated the link between the amount of details present in the image known as visual complexity and human memory, reporting that high complexity reduces the ability to focus on important visual information. Forsythe [45] examined different measures of complexity and the extent to which they may be compromised by a familiarity bias. In [46] researchers were trying to understand the association between color and visual complexity in abstract images, knowing that color is one of the most important aspects of vision and that it elicits both aesthetic and psychological responses. They discovered that images evaluated as visually complex lead to harder differentiation between hue colors, one of the color appearance parameters.

In our work, we are advancing the state of the art when it comes to computational models for studying memorability, by analyzing the role of the image style. In photography, style is interpreted by the photographer in the way they create and capture an image, i.e., the way they use lenses, lighting, filters, the composition of the image and through various techniques for processing the image after it has been taken. In visual arts, style is the manner in which the artist depicts, interprets and expresses his or her vision and it frequently permits the grouping of works into related categories. Thus, artists are trying to develop their own unique drawing and instantly recognizable artistic styles (Fig. 2.1.).



Modernism *Gloria G. Bernstein*

Impressionism *Claude Monet*

Abstract art *Wassily Kandinsky*

Cubism *Pablo Picasso*

Abstract *Piet Mondrian*

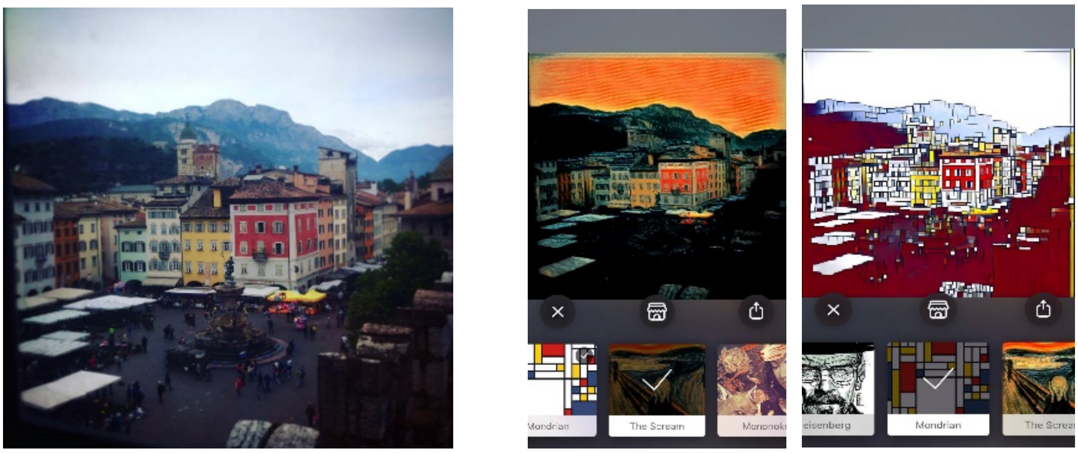
Expressionism *Edvard Munch*

Figure 2.1: *Example of different types of art painting styles*

Expanding knowledge about all the differences in painting styles becomes truly fascinating, especially when placed in the context of computer vision. Despite all the complex calculations and tasks in which AI managed to outperform humans, the potential of unlocking the computational creativity, that is modelling and replicating creativity of humans and using AI to generate new ideas and some novel combinations has been an open challenge for a long time. Today we can confirm that art with AI is what enables and facilitates photographs to be turned into artworks. Moreover, AI began to generate aesthetically attractive music and poetry too. As Moruzzi stated in her research on Creative AI [47], further questions about the creativity and intentionality exhibited by AI will emerge.

Since our focus is on understanding memorability, while considering the importance of image style, we should explain what style, specific to Neural Style Transfer, really means in the context of this work. It refers to the texture of an image which captures the geometric shapes, all the patterns, the colors, the brush strokes in paintings or the way painters apply a brush or palette knife. Image style is typically difficult to copy and reproduce manually. Another concept important for the research presented here is the content of an image and refers to the specific layout and positioning of the objects in one image.

An optimization technique called Neural Style Transfer refers to the combination of the two: it combines the content of an image with the style of a different image. As a result, this technique is effectively transferring the style of an image and it is capable of generating fascinating results that are difficult to produce manually (Fig. 2.2). In our case, the generated image has the identical content, while the style of an image is a property that varies. To the best of our knowledge, no works so far have investigated the problem of transferring a style to an image in order to enhance its memorability.



a) original content

b) stylized images with list of filters in the bottom

Figure 2.2: *An example of photo-editing mobile application Prisma that uses Neural Style Transfer*

In addition to image style, we decided to investigate the connection and relationship between memorability and other perceptual attributes, which we perceived relevant to our work, such as interestingness. Human interest in photos was examined through a series of psychological experiments Gygli et al. [41] conducted, trying to explain how it relates to the extensive list of image attributes they explored including emotional, aesthetic, content-related aspect and memorability. The existence of a strong correlation between aesthetics and interestingness has been confirmed, while the expected correlation of memorability with interestingness has not been proved. This was important to verify as intuition can often be misleading, like in case of Isola et al. [10], who showed that human prediction of what is memorable, or how the researchers defined the assumed memorability, was negatively correlated with actual image memorability. The development of methods for objective measuring this image property, as a basis for any further quantitative analysis, could help overcome the lack of consensus among researchers on this issue.

2.2.2. Marketing perspective

Memorability is one of the terms frequently used in digital marketing to describe the ability to memorize the processes and features offered by a website or application after leaving it and then returning to it after a time. It is also one of the principles of usability [48], the level of simplicity and accessibility of use, referring to what extent an application has an attractive design and how intuitively it can be used by a new user, how easy a user interacts with the application or any other product. Memorability is particularly important in the user experience design process (UX), considering that a user needs some time to learn how to navigate a website or any application and find what he searches for. Adjusting, customizing and advancing the usability and accessibility encourages users to navigate easily and to interact and engage effortlessly with the product they use. Memorability indicates how smooth and fluid is the users' navigation after a period of not using it; they should be able to remember and evoke how to do something when they come back after a long time. High memorability is a measure of user engagement. If it is natural, logical and intuitive, people will keep coming back. For instance, we all have certain expectations when we swipe across the mobile screen, in other words, we all have a mental model of how a certain user interface works and designers have a task to facilitate mental model generation so that people can recall easily. These models serve also to prevent users from having to investigate things from scratch with every new experience, which is why they belong to the most important concepts in human-computer interaction (HCI).

This concept has been largely explored by the entrepreneurs and marketers in the context of their brand memorability improvements. Mapping those specific elements able to leave long-lasting memories should strengthen positive user experience and serve the purpose of advertising. At first, visibility has been discovered to be the most popular indicator for a digital campaign's success [49]. However, visibility was not able to capture whether an ad had a lasting effect on the consumer or not. Thus, the list of KPIs (Key Performance Indicators) of a successful marketing campaign had to be revised.

As a result of unifying marketing, graphic design and engineering, many attractive digital tools have been developed. An example of one such application is Instagram, today recognized as a photo and video sharing social network service owned by Facebook, whose primary focus was solely on communication through images. Stylish, modernized and pivoted version of the app simplifies the adjustments and transformation of visual content, simultaneously involving the end-user in the media modification and organization process, with filters and hashtags, respectively. The goal is to create visually more appealing images and videos, according to single, individual and predominantly subjective criteria for the content "likeability". In addition, all posts can be shared publicly or with pre-approved followers, users can also browse other users and- they can like each other's content. This also raised diverse scientific interests.

Some studies started concentrating on understanding and predicting likeability [50] and virality [51] of images, rather than their memorability.

Typically, products based on the philosophy of image filtering are used for modifying an image to emphasize and highlight certain features, to remove other features, to adjust the intensity of the effects, to smooth, sharpen or set up a new color. However, once the editing process is complete, we cannot claim with certainty that the chances of remembering that photo will be increased, which implies a completely different purpose of further use of the given image.

By identifying human behavior patterns, along with understanding their habits and perception more precisely, we managed to extract some knowledge useful for a range of other applications, such as how to use visual elements in study and training materials [52]. This further contributes to a better understanding of factors disrupting the process of learning and memory, such as the effect of complex material which causes cognitive saturation. Through its scientific approach to the design of learning materials, cognitive psychology has differentiated types of cognitive load and articulated them through the Cognitive Load Theory (CLT). This theory explains how human attention is divided between multiple sources of visual information, which adds to the cognitive load, making it more difficult to be memorized [53]. Similarly, once an image observer reaches his or her cognitive capacity, they become saturated which makes it problematic or completely impossible for the brain to further process information. Researchers in [54] were concentrated on memory encoding to determine material-specific differences in brain activity patterns. All these findings allow us to understand parts of images and study material that could be easier to memorize, as well as the segments of a single video which have the greatest memorability potential. As an outcome, a recent study demonstrated that users are more likely to watch videos which have highly memorable and interesting video summaries [55].

A positive correlation between memorability and popularity has also been confirmed [12], meaning that memorable images have a higher chance of becoming more popular, which is another interesting discovery for further marketing and brand development implementations. Subsequent studies showed that it is possible to predict the number of views that an image will receive on social media even before it is uploaded [56]. Other studies lead to the conclusion that high-level concepts like the presence of faces are what contribute mostly to memorability [39], which helped marketers implement those findings in their future campaigns, connecting this way with their audience on a human-level, while being more distinctive from the competitors and easier to remember. The purpose of putting a human face picture in the billboard, or any other space for image advertising is surely to draw observers' attention, while the rest of the space is used for communicating the message, or simply applying the logo, depending on the previously achieved level of brand recognition. These findings suggest that high-level semantic attributes and features are an efficient way of characterizing the memorability of photographs. The explanation provided reveals why certain photos are easy to remember, while others are quickly forgotten. However, the particular work does not exploit the flexibility and potential of optimization techniques in modifying an image in order to enhance specific high-level attributes, like memorability. Our paper proposes an innovative approach to automatically increase the memorability of an arbitrary input image by changing its style, in other words, by modifying its low-level features while preserving the high-level content.

One of the advanced advertising strategies replaces randomly selected human faces of strangers with celebrities (Fig. 2.3), which undoubtedly contributes to better brand recognition. Celebrity-driven ad campaigns, known as celebrity branding or celebrity endorsement are perceived as powerful mechanisms for brands to become more memorable.

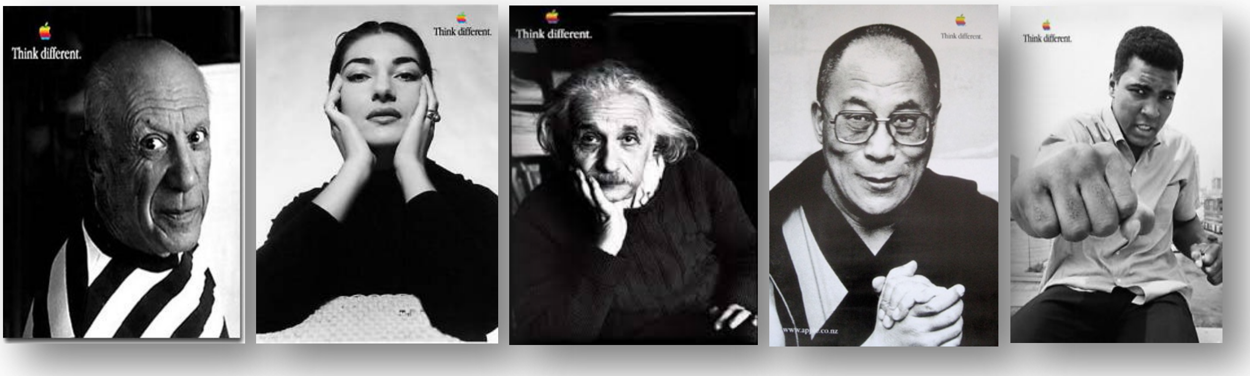


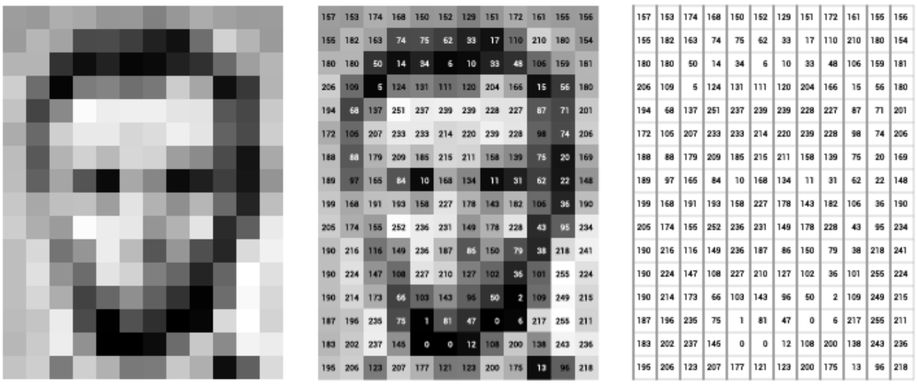
Figure 2.3: *Iconic 20-th century personalities used in Apple's "Think Different" campaign between 1997-2002*

With the development of AI agents, campaign materials are modified in real time, which maximizes consumer interest and message retention. Together with exploring the possibilities of optimizing images for ad use, technologies have been developed and used by marketers to customize their promotions [57], for example Dynamic Creative Optimization (DCO). This display ad technology creates personalized ads and ensures that continuously change according to each shopper's preferences and browsing history (i.e., geolocation, similar products seen, time of display etc.). Motorized by machine learning algorithms these forms of programmatic advertising are helping marketers to undertake their activities using real-time technology and decrease the uncertainty of their advertising campaigns. Using the correct format of the ad in the right environment is what makes them viewable. However, viewability does not always translate to memorability.

2.3. Computer Vision and Image Memorability

Giving machines a sense of vision humans possess essentially means finding a way of translating those human senses to a language understandable to computers. We employ our visual system for things like navigation, the way we recognize and pick-up objects, in the reception of light, but also in complex behaviors and human emotions. As visual information passes through the visual hierarchy, the complexity of the neural representations increases.

Computers are able to process images, or more precisely image pixels, by essentially translating every pixel they observe in the image into one single number. This way we can represent a grayscale image as a two-dimensional matrix of numbers, one for each pixel in that image (Fig. 2.4).



157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

Figure 2.4: *Image representation by the two-dimensional matrix of numbers*

However, for our research we need also to consider the color images or the RGB images, meaning that we have to represent them as three of these 2D images stacked on top of each other for each color channel, so one for the Red, one for the Green and one for the Blue.

Typically, our motivation is to try to detect the presence of certain features in the image and to try to extract those that we need for further analysis. To do this in a hierarchical fashion we need neural networks, so that we can both learn the visual data and the hierarchy of those features in order to learn those visual features.

Studies of memorability and the quality of being worth remembering relates primarily to AI-subdiscipline of Computer Vision (CV) whose complex processes are often described by analogy with human vision. This interdisciplinary scientific field investigates various computer procedures used to obtain high-level understanding from digital images or videos. From the point of understanding the human visual system, computer vision aims at developing computational models which will automate tasks that the human visual system can do. Our visual cortex is fascinating on so many levels and responsible for processing and interpreting visual data which will give certain perception and formulate our memories. It helps us understand a lot from a very little information. For example, we can tell the whole story behind blurred photographic scenes by utilizing that hardly visible context and connecting it with our prior knowledge. From the engineering point of view, computer vision aims to build autonomous systems which will perform some of the tasks [58] the human visual system can perform. This interdisciplinary scientific field enables the information extraction from digital images or videos [59], continuously exploring innovative ways to automate the most complex tasks.

Naturally, the development of computer vision and progress in that domain would have been impossible without prior progress in other fields, such as neurobiology. One significant period in our history happened in the early 1960s at Harvard University after the breakthrough discoveries about the visual system and visual processing of two scientists David Hubel and Torsten Wiesel, honored by a Nobel Prize [60]. These researchers managed to record electrical activity from individual neurons in the brains of cats and noted that specific patterns showed in a slide projector, stimulated activity in specific parts of the brain. Inspired by this, a neuroscientist at MIT - Massachusetts Institute of Technology David Marr [61] managed to integrate the results from psychology, AI and neurophysiology and create new models of visual processing. In other words, he started formulating computer vision to mimic human vision capabilities. These are only some of the examples of the early beginnings of how we started solving object classification, recognition and detection problems. Over time researchers managed to design more sophisticated algorithms to organize, annotate but also to retrieve multimedia data.

Cambridge University Press Machine Vision Textbook written by the S. J. D. Prince [62] recognizes Computer Vision as a subfield of AI and Machine Learning (Fig. 2.5).

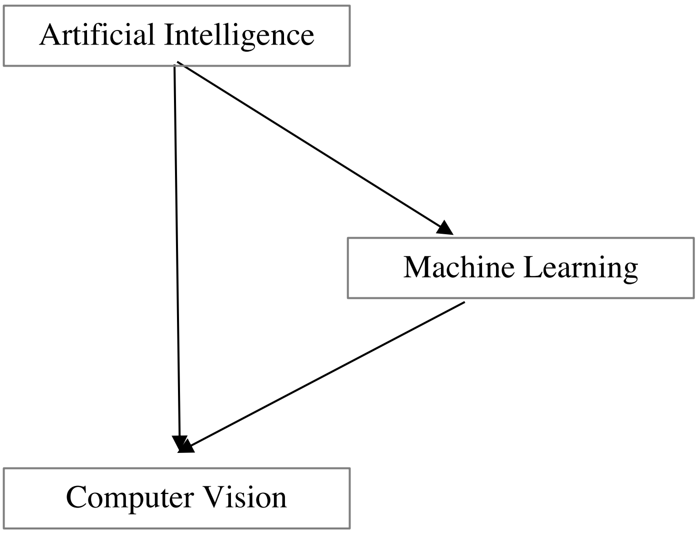


Figure 2.5. *An overview of the Relationship of Artificial Intelligence and Computer Vision*

State of the art computer vision techniques have generated a lot of interest in solving various challenges in different fields of research and industry, from medicine, gaming industry or special effects in movies, to applications in automotive industry, some other tasks such as face detection or object recognition, available in small devices like cell phones and so, adjusted to accommodate individual user's visual needs. In the last couple of years, the AI community has offered a number of industry-ready solutions and some of the techniques demonstrated incredible advancements in performance, outperforming huma-level achievements.

Concrete examples for this are coming from image and object recognition and classification tasks, where state-of-the-art networks trained on a database of labeled images were able to classify objects better than a human. They were developed by Andrej Karpathy[3] et al., who trained on the same task for over 100 hours. Back then a PhD researcher, today the director of Artificial Intelligence and Autopilot Vision at Tesla, Karpathy was only one of the participants in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) back in 2014, one of the largest challenges in Computer Vision. The challenge is held annually and brings together teams to compete and claim the state-of-the-art performance on the dataset, based on a subset of the ImageNet dataset, collected first time in 2009. World's leading automotive company Tesla is continuously working on transition to sustainable energy with electric cars and is very close to having fully self-driving vehicles, a challenging process in which computer vision and image understanding have a largely significant role.

Although there are various computer vision applications, we will focus on the most relevant ones for the task we are solving in this paper:

- ***Image Stylization*** – refers to a style transfer or a Neural Style Transfer (NST), the task of learning style from one or more images and applying that style a new image, like in the example of applying the style of famous artwork to new photographs described in [63]; prior to NST, this image style transfer was performed using machine learning techniques based on image analogy [64].

- ***Image Synthesis*** – this refers to the ability to produce images starting from images or another type of information i.e., random noise, text describing the image or a feature of the image. It is a task of generating targeted modifications of existing images or entirely new ones by changing the style of an object in a scene or adding an object to a scene.

Before understanding details of both approaches, it is important to emphasize one significant function they have in common, truly relevant to our work and called ***image-to-image-translation***. It is defined as translating the possible representation of one scene into another, for instance mapping a grayscale image to RGB, or generating an image from the edges only. This approach takes images as a conditional input, which was exactly our approach towards

[3] https://karpathy.ai/

building a deep architecture for generating a memorable picture from a given input image and a style image as demonstrated in the Fig. 2.6.

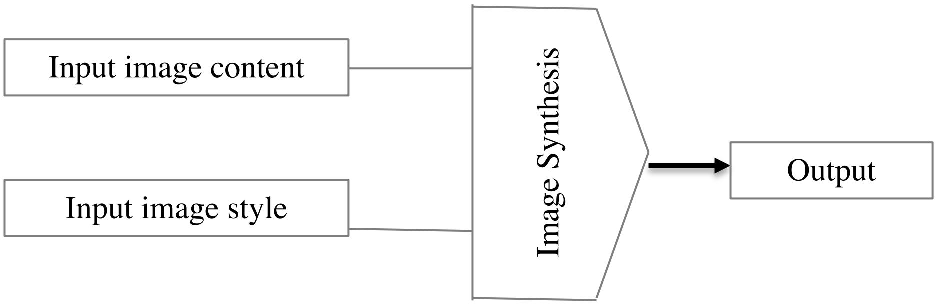


Figure 2.6: *Block diagram of the image synthesis process in image-to-image translation problem*

The goal of image-to-image translation is to learn the mapping between an input image and an output image (Fig. 2.7) and it has a range of applications i.e., style transfer, photo enhancement and more.

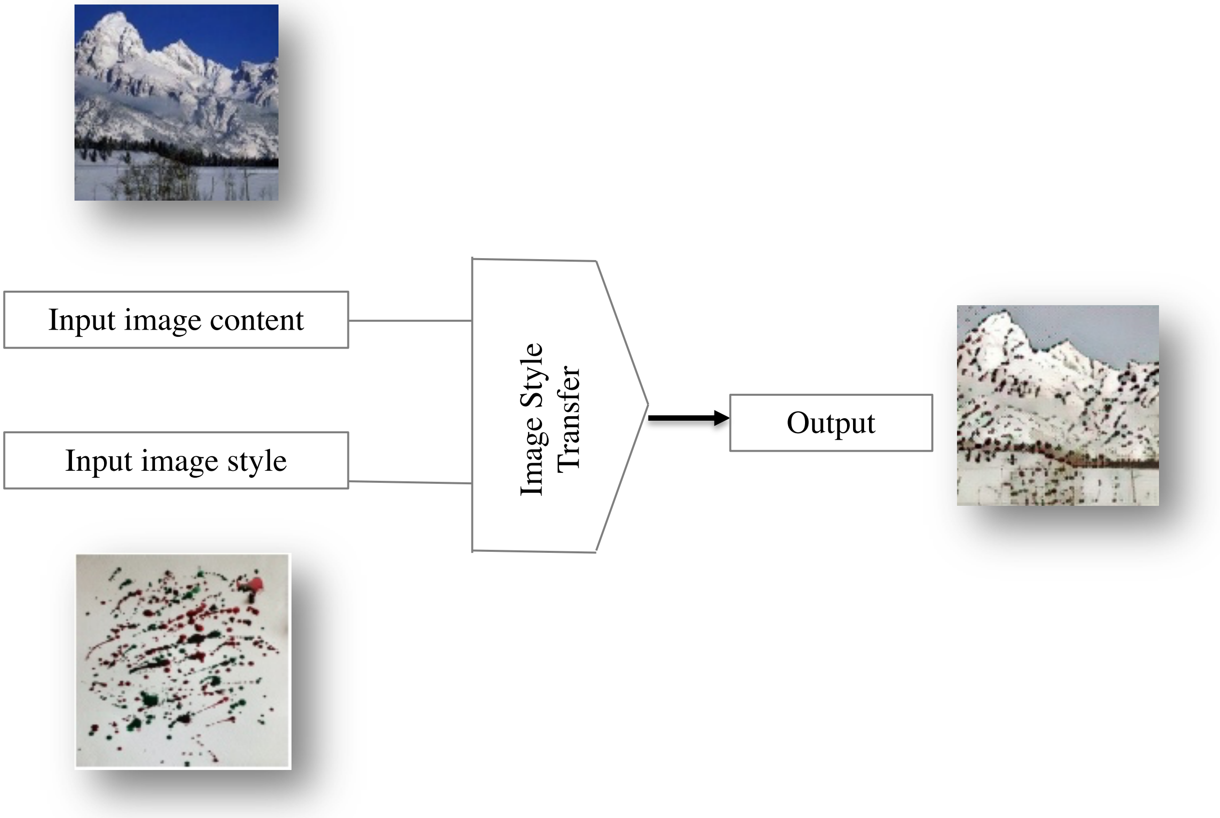


Fig. 2.7: *Block diagram of the image style transfer in an image-to-image translation problem*

One task that is directly related to image-to-image translation is style transfer, where the goal is to transfer the style of one image to another, while maintaining the content of the input, exactly as defined within the scope of our research.

Another example of image-to-image translation is super resolution where the resolution of the input image is enhanced.

It should be noted that due to the absence of a clear borderline between image-to-image translation and style transfer, it is not unusual to see the two concepts being used as synonyms in different publications, especially if the translation is performed between two different domains.

Image Stylization concerns transferring style of a reference photo to a content photo with the constraint that the stylized photo remains photorealistic. This means that the reproduction attempt is performed as realistically as possible, which is typically a very challenging task due to the tendency of existing stylization methods to generate noticeable artifacts. An image artifact is any feature that appears in the output image and has not been present in the original one. The ability to identify the source of an artifact is related to our understanding of the original material, which is particularly important in medical image processing, as the artifacts can be mistaken for the pathology. There have been methods developed and proposed to address these issues. The method in [65] consists of a stylization step which transfers the style of the reference photo to the content photo and a smoothing step which ensures consistent stylizations.

Artistic stylization of images is studied in computer graphics under the label of non-photorealistic rendering, the means of generating images that do not aspire to realism [66]. It refers also to creating an illusion of reality in the mind of the observer by combining computer graphics with artistic techniques. The necessity of providing forms of visualization which are non-photorealistic arises in various market segments. For instance, attracting people with visually appealing design of houses or even only kitchens which typically requires one designer or an artist who will create a stylized depiction of the design. In contrast, this allows an assembly of predefined 3D objects to be designed very quickly and accurately, which can then be provided to a rendering system from which photorealistic images can be generated.

Despite the fact that computer-generated image content has been a concept since the mid-twentieth-century, the growth of interest and research curiosity on *Image Synthesis* continued its expansion conditioned by progress and wide application opportunities in domains, for instance in computer games, movie industry, advertising industry, virtual reality and various social media platforms, on top of medicine, science and engineering. Consequently, numerous methods have emerged so far. Long before the deep learning revolution that started a decade ago, the preconditions for implementations in that domain date back to thinking machines by Turing and other concepts fundamental to deep learning, i.e., Recurrent Neural Networks, back-propagation, Reinforcement Learning and Convolutional Neural Networks. Since then, there has been a rapid development on increasingly more difficult problems and image synthesis has improved dramatically providing the means to efficiently generate large quantities of training data while controlling the generation process to provide the best distribution and content variety [65]. The goal of this process of creating new images is to respond to the deep learning application demands where synthetic data has the potential of becoming a vital component in the training pipeline. Synthetic images are commonly used to verify the correctness of operations by applying them to known images. Image Synthesis or rendering in computer graphics means the process by which a virtual or geometric scene is converted into an image and its methods can be used to create a stand-alone training dataset for machine learning optimization. The goal is to create synthetic images that can be mistaken for photographs. This process can be performed in different ways, with various approximations which can affect the computational complexity and accuracy.

There are various problems related to image-to-image translation and several approaches trying to solve those different challenges using both supervised and unsupervised methods have been proposed. Machine learning tasks are usually divided into three broad categories, based on the nature of the data:

- *Supervised learning* is the machine learning task of learning a function that maps an input to an output based on example input-output pairs [14], or in other words, prediction of the right answers is based on the data we gave to the computer to learn from i.e., training data. Each example is a pair consisting of an input object and a desired output value (also called supervisory signal). In this approach, the model infers a function from supervised or labeled data, which is how it got the name. Supervised learning deals with two types of problems, namely classification and regression ones. Yet, its major limitation is exactly the fact that it must use paired images as supervision.

- *Unsupervised learning* allows us to approach the problem with little or no idea about the result, since there is not available feedback based on the prediction. This type of a machine learning algorithm is also known as self-organized learning in contrast to supervised which usually makes use of human-labeled data [67]. Due to the fact that no labeled dataset is provided, the model can either learn from the existing information according to which it can distinguish between the categories. Otherwise, in case this information is missing or in case the model is not able to do this correctly, it follows *backward propagation* for reconsidering the image. Back-propagation is the essence of neural net training, especially in fine-tuning the weights of neural net, based on the error rate (i.e., loss) obtained in the previous iteration.

2.3.1. Image Manipulation

While Computer Vision aims at understanding an image and what specific image represents, Image Processing comes in the form of Image Manipulation. Recent works have shown that it is possible to manipulate images with the ultimate goal of altering perceptual properties like aesthetic values [68], with a neural network designed so that it has two branches for predicting attention and analyzing aesthetics and evoked emotions [65] by exploring new aspects of photo and human emotions through psychovisual studies and finally – memorability [66] of face photographs.

Wang et al. [68] from Beijing Lab of Intelligent Information Technology evaluated different ways of adjusting and modifying perceptual properties of an image by defining the photo cropping problem as a cascade of attention box regression and aesthetic quality classification, based on deep learning. Basically, they examined the way people determine the parts of a photograph they consider adequate and appropriate for a crop of a certain image after which they proposed a deep learning-based photo cropping approach which is driven by human attention box prediction and aesthetic assessment. The role of exploiting the attention information was to avoid discarding important information, while the aesthetic assessment was employed to ensure the high aesthetic value of cropping results, with the goal of imitating human interpretation of the beauty of images. Early visual attention prediction models based on various low-level features like color, intensity and other are being advanced with approaches that employ also high-level features, either from person or face detectors (learned from specific computer vision tasks). The most recent models, driven by the success of deep learning in object recognition, are models based on attention and they commonly give very impressive results. In general, these methods are categorized as attention-based (with an aim of preserving the main subject and visually important area) or aesthetic-based models (emphasizing general "attractiveness" of the cropped image). In their approach, Wang et al. decided to predict an attention bounding box, which covers the most informative regions of the image. Finally, these researchers evaluated the whole cropping model on two publicly available image cropping datasets and achieved superior performance in comparison to the state-of-the-art, regardless of the limited availability of training data for photo cropping.

Understanding that an image is able to evoke a deeply strong emotion in the observer, psychovisual studies were exploring new aspects of photos and human emotions. Knowing that by using different filters and photographing techniques the same exact image can elicit largely opposite emotions, researchers were fascinated and strongly motivated to mimic this process and understand its details. Peng et al. [65] showed that different people have different emotional reactions to the same image, which was a completely new discovery, opposite to the previous ideas about a single dominant emotion for each image. Their studies showed also that the same person may have multiple emotional reactions to one image. And so, they stress out that predicting emotions in “distributions” instead of single dominant emotions is important for many applications. Furthermore, they showed that not only can we change the emotions initially evoked - by adjusting the tone of the color and texture related features, but more importantly that we can also choose in which “emotional direction” this change will go, by selecting a target image. And so this work has introduced the idea of representing the emotional responses of observers to an image as a distribution of emotions, while they also described methods for estimating that emotion distribution for an image and finally a method for modifying an image so that it pushes the emotions evoked towards a target image.

Knowing that face photographs are expected to be remembered, either because of personal relevance, images commercial interests or because pictures were intentionally and consciously designed to be memorable, Khosla et al. [66] decided to examine the possibility of making a portrait more memorable or more forgettable in an automatic way. Researchers provided a method to modify memorability of individual face photographs, while keeping the identity and other facial traits of the individual fixed, like the age, attractiveness and emotional magnitude. In a crowd-sourcing experiment and with an accuracy of 74%, they showed that face photographs manipulated to be more memorable or more forgettable are indeed more often remembered or forgotten.

Evaluating, quantifying and modifying the memorability of faces has already found broad-spectrum implementations within computer vision and graphics, one of them is certainly mnemonic aids for learning – formula or rhyme used as an aid in remembering, strategies designed to help students improve their memory of important information by connecting new learning to prior knowledge through the use of visual (and acoustic) cues; other applications are in photo editing apps for social networks and other various tools for designing memorable advertisements.

As one of the most enigmatic properties of an image, emotiveness, defined as the ability to trigger an emotional reaction in the viewer [65], has been also studied with an aim of expanding empirical evidence about the role emotions play in the image retrieval process. In our part of the emotion research process, while working on this study, we have applied the most recent discoveries in the affective science, the study of emotions or affects, in order to classify emotions and understand the emotion elicitation, emotional experience and the recognition of emotions. While exploring the dimensional models of emotions in modern psychology, we discovered one thing all these models had in common: the attempt to conceptualize human emotions by defining their exact position in two or three dimensions. Furthermore, most dimensional models incorporate valence and arousal or intensity dimensions, which is why we found a number of papers proposing novel approaches in this direction.

A methodology [69] based on semantic-segmentation was proposed to modify the valance-arousal score of images with natural scenes. Khosla et al. [12] showed that by removing visual details from an image through a cartoonization process (i.e., cartooning), its memorability score can be modified. However, these researchers did not provide a methodology to systematically increase the memorability of pictures. The same group [66] demonstrated also that it is possible to increase the memorability of faces, while maintaining the identity of the person and properties like age, attractiveness and facial expression.

Yet, to the best of our knowledge, our work is the first to attempt to automatically increase the memorability of generic images, not only faces. We cast the problem of increasing image memorability as a problem of selecting the most “memorabilizable” style filters, thus, addressing the task of image manipulation under the popular “editing-by-applying-a-filter” paradigm. There are a number of these easy-to-use apps that contain multiple filters which are helping users improve and get the most out of their images, but none of the existing apps enables automatic memorability increase of an image a user wishes to edit, adjust and manipulate before posting in publicly available platforms. Users are exposed to image editing tools and various easy-to-use filters advertised as mechanisms which will make their photos and videos more “visually appealing”, not more memorable.

In a similar line of though, researchers wondered how to accurately predict which images will be remembered and which will not. Recent experiments showed near-human performance in estimating, measuring and predicting visual memorability, when MemNet, a model trained on the largest annotated image memorability dataset, LaMem, is concerned. Details will be presented in the next Chapter.

2.3.2. Neural Style Transfer

Style Transfer is a computer vision technique that enables recomposing the content of an image in the style of another. It is a particular application of image-to-image translation where the goal is not to translate from one image domain to another, but to create a new picture by “merging” the two different images, namely content and style image. What Style Transfer practically does is that it takes two images, a content image and a style reference image, and it blends them together so that the output image preserves the fundamental elements of the content image, while appearing stylized with the style reference image (in the style of that image). Style Transfer is an example of Image Stylization, an image processing and manipulation technique that has been studied within the field of non-photorealistic rendering. Some examples of Style Transfer images are shown in the picture below (Fig. 2.8).

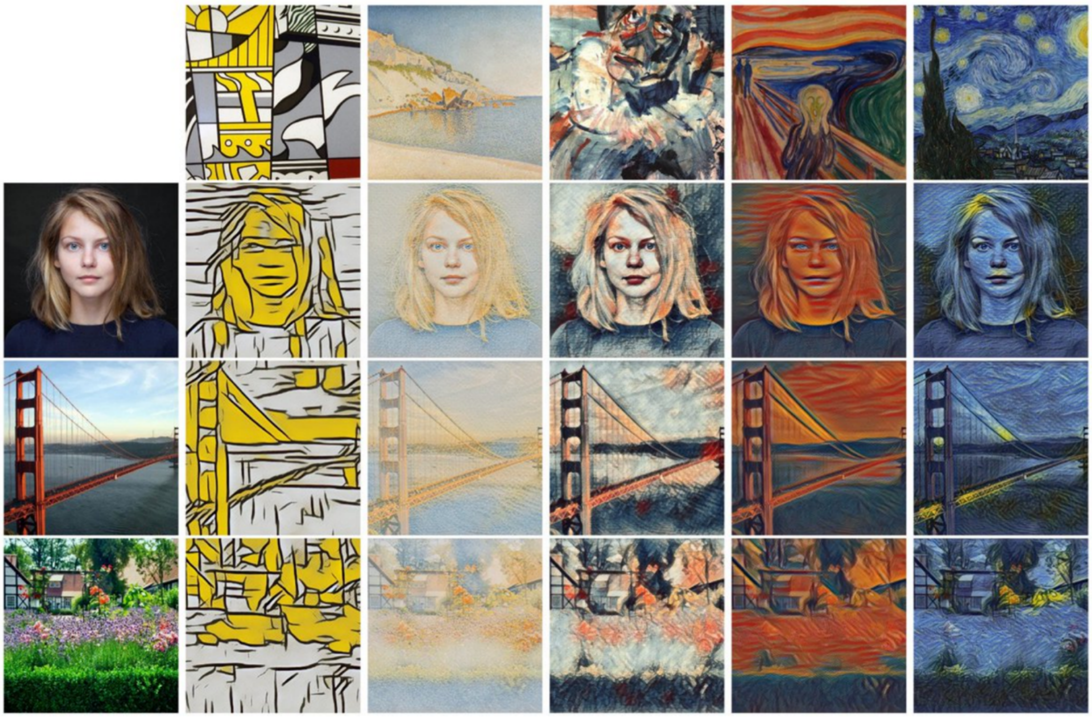
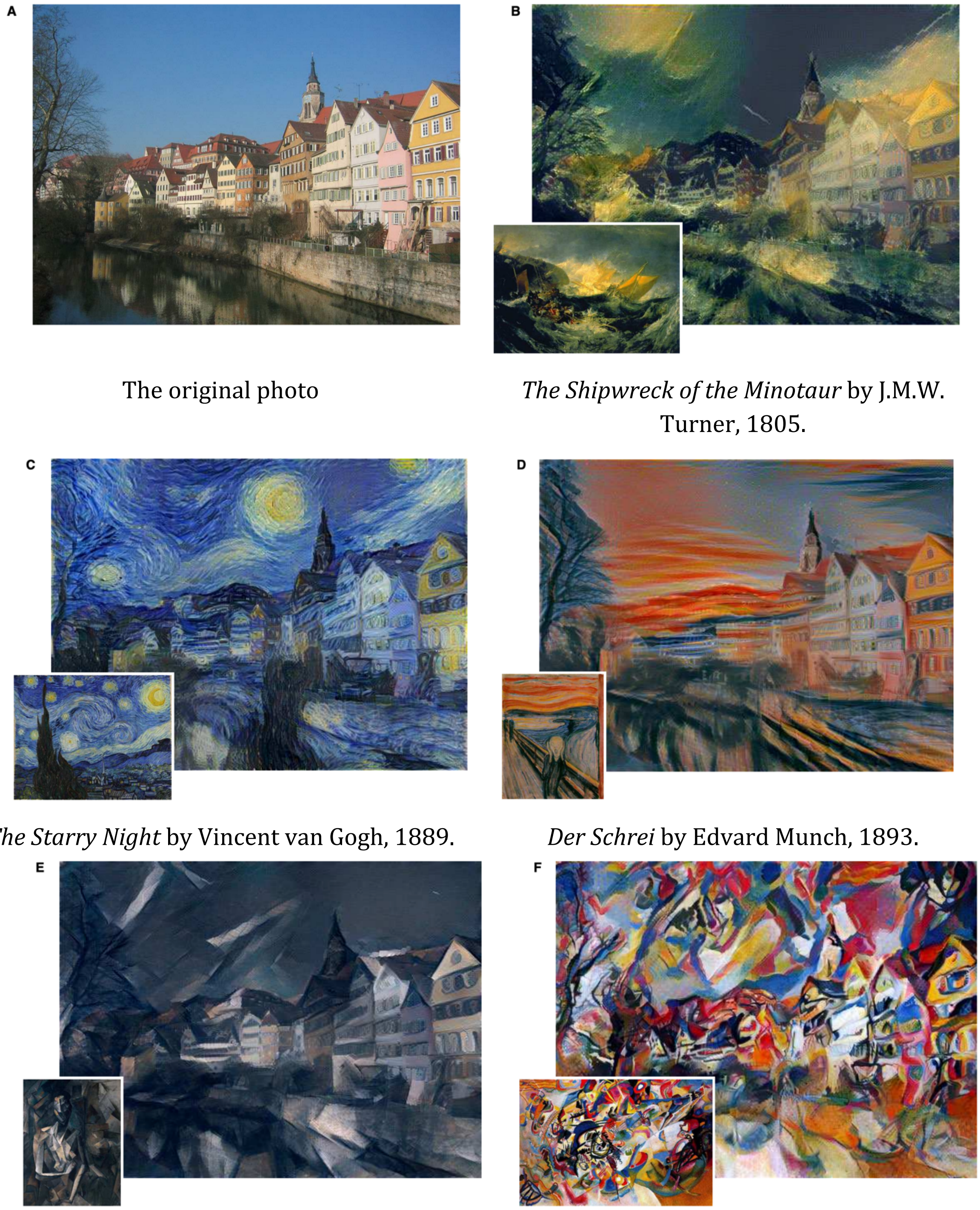


Figure 2.8: *Different examples of Style Transfer. Images in* [70]

However, if we think of solving the challenge of transferring style between images in a traditional way, with a supervised learning approach, then a learned Style Transfer would have required a pair of input images, namely one original image and an artistic representation of that original image and from there a machine learning model learns the transformation and is able to apply it to new images. From a practical perspective, this approach does not have many advantages, caused by the absence of all those image pairs, and so the new path has been created, called the Neural Style Transfer (NST).

An image style is defined typically as a texture. It is similar in the whole picture, in its different parts and it is repeated across the image, while the content is composed of local features which represent the shapes of the objects in those images. Here we can explain the goal of the Style Transfer and that is to mix the two different features together.

An excellent example of image Neural Style Transfer was offered by a group of authors led by Leon A. Gatys [63] who tried to imagine the way one photo might look like if it was painted by a famous artist like Vincent van Gogh, Edvard Munch, Pablo Picasso or Wassily Kandinsky. Indeed, this is the first example of the style transfer that was able to produce impressive results. These researchers decided to combine the content of a photograph with the style of several well-known artworks by employing deep neural networks that can turn it into a reality and allow these transformations. The images were created by finding an image that simultaneously matches the content representation of the photograph and the style representation of the artwork (Fig. 2.9).



The original photo

The Shipwreck of the Minotaur by J.M.W. Turner, 1805.

The Starry Night by Vincent van Gogh, 1889.

Der Schrei by Edvard Munch, 1893.

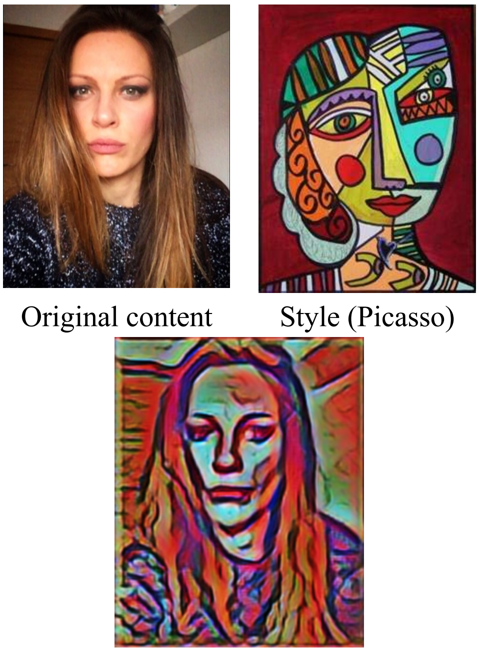
Femme nue assise by Pablo Picasso, 1910.

Composition VII by Wassily Kandinsky, 1913.

Figure 2.9: [63] *The original photograph depicting the Neckarfront in Tubingen, Germany is shown in "A" (Photo: Andreas Praefcke). The painting that provided the style for the respective generated image is shown in the bottom left corner of each panel.*

Essentially what Gatys et al. [63] achieved is an impressive result of a difficult image processing task, particularly rendering the semantic content of an image in different styles. Arguably, a major limiting factor for previous approaches has been the lack of image representations that explicitly represent semantic information and therefore allow to separate image content from style. Transferring style from one image to another is considered to be a problem of texture transfer, where the goal is to synthesize a texture from a source image, while constraining the texture synthesis in order to preserve the semantic content of a target image. Most previous texture transfer algorithms [71, 70] that managed to achieve remarkable results still struggled with the same limitation – all of them used only low-level image features of the target image to inform the texture transfer. The fundamental contribution this research provided is that they managed to find image representations that independently model variations in the semantic image content and the style. These representations derived from convolutional neural networks, optimized for object recognition, which makes high-level image information explicit.

Neural Style Transfer therefore facilitates separation and allows recombination of the content of one image with the style of another, enabling this way the production of new images. In practice: a selfie would supply the content and the Picasso's painting would be the style reference image, so the output result would be our self-portrait that looks like "Picasso's original" (Fig. 2.10).


Original content

Style (Picasso)

Output image

Fig. 2.10: *Image modification with Prisma application*

Not exactly. These limitations are caused by imprecise definitions of what exactly illustrates the style of an image i.e., the brush stroke in a painting, the color map, some shapes, or a composition of a scene in the stylized image, or even the subject of the image. Consequently, it is not clear if the image content and style can be separated completely in the first place. Style Transfer implementations are limited to a pre-selected list of styles, due to the requirement that a separate neural network must be trained for each one of them.

Despite being defined as flexible, since it was able to work with any content or style image, the proposed algorithm was defined as “expensive”, because it requires an optimization phase for any run.

Nevertheless, these findings are still inspiring considering the way the neural system automatically learns image representations that allow (at least to a certain extent) the separation of the two, the style of the image from the content of image.

These advances are also allowing almost anyone to enjoy the process of creating an artistic image and some users also to make business out of it, not only by creating popular applications to simplify the utilization of these transformations but also to sell the AI artworks. In 2018 Christie’s became the first auction house to offer a work of art created by an algorithm. The portrait in its gilt frame depicting a portly gentleman, created fully by an AI algorithm, has been sold for $432,500 (Fig. 2.11). This fact could serve as an additional empowerment to all people to try out the existing apps and explore their own creativity, while playing with Style Transfer.

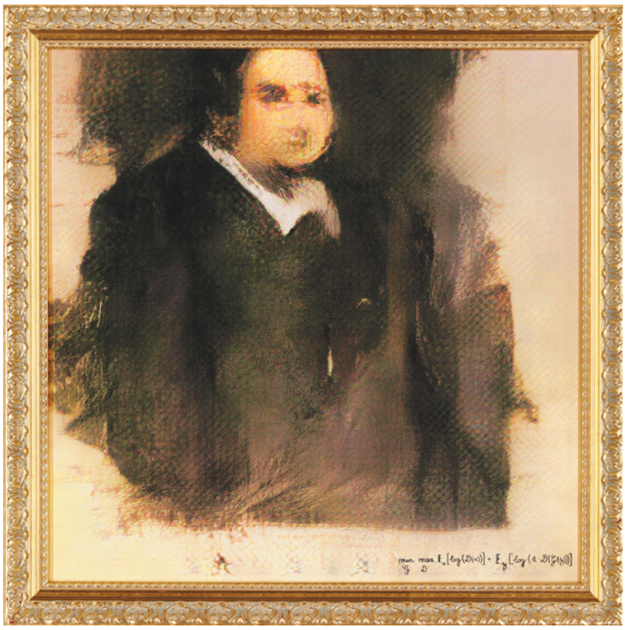


Figure 2.11: *A portrait produced by Artificial Intelligence*

Previous decade has shown some revolutionary advances in Computer Vision. By using Deep Neural Networks (DNN), this area offered not only the most accurate object detection and image segmentation methods, but it also advanced the prediction of the outcome of various events, and consequently modified the human-computer interaction (HCI). An example for this is Prisma, a photo-editing mobile application that uses neural networks and artificial intelligence to apply artistic effects to transform images. Prizma's art filters allow their users to apply the style of the most famous painters, such as Picasso or Salvador Dali to any photo. User interactivity with Prizma leads to the transformation of random photography into a real work of art, similar to original style of these famous artists. Amazed by the interesting interplay between content and style of an image and even more interesting outcomes offered by this application, it achieved a record number of downloads by 7.5 million users only one week after its launch [70]. This application was inspired exactly by the DNN Computer Vision models, where a computer system uses neural representations to separate and recombine content and style of arbitrary images, providing a neural algorithm for the creation of artistic images [63].

Following the development and the evolution of Style Transfer research, from single and multiple style models [72] to arbitrary-style approaches, it becomes clearer what the possibilities are and what is the trend of further development and Style Transfer employment. Training one such model requires two networks, one pre-trained feature and a transfer network. In this process, some layers learn to extract the image content (i.e., shapes or positions of the objects in the images), while others learn to focus on texture and patterns (i.e., brush strokes of a painter). Essentially, by running the two images through a pre-trained neural network, Style Transfer is comparing the similarity between the pre-train network's output at multiple layers. Those images that provide similar outputs at one layer of the pre-trained model are expected to have similar content and at this stage we can easily compare the content and the style of two images but we are still not able to create an output or the stylized image. This happens in the next stage when the transfer network or the image translation network, takes one image as input and outputs another image. Typically, style transfer networks have an encode-decoder architecture.

Training starts with running several style images through the pre-trained feature extractor and the outputs at various style layers are saved for the later comparison. The same procedure applies to each content image, they pass through the pre-trained feature extractor where outputs at various content layers are saved likewise. The content image then passes through the transfer network and it outputs a stylized image. Running through the feature extractor itself, the stylized image then outputs both the content and style layers.

During this process, our goal is to optimize the complete transformation action and to make the most effective use of a resource we have. When using an arbitrary image as an input, like in our case, this part is particularly challenging and the goal is to make the decision which will help us minimize the loss function, the function that maps an event or values of the variables into a real number [73], intuitively representing some “cost” associated with the event. This translates into the quality of the stylized image after the transformation process, which is exactly defined by a loss function; both the extracted content and style features of the stylized image are compared to the original content image and the reference style image. In this process, the weights of the pre-trained feature extractor remain fixed. However, if we decide to change the weights, we will achieve different stylization in the output image, which will have a direct influence on the quality of this image.

There are too many possibilities for building neural network architecture. Yet, if we are to define several directions which helped us in our research, we should start with:

- *Single style per model,* in research conducted by Johnson et al. in 2016 [74] who were the first to train an independent neural network to stylize images in a feed-forward (single) pass. These researchers were considering the image transformation problems by combining the benefits from all previous approaches and training a feed-forward network for image transformation and real-time optimization tasks. In their approach, a single transfer network is trained for each desired style.

- *Multiple styles per model* – an option for blending more than one style together was published a year later by Dumoulin et al. in [25], when researchers investigated the construction of a deep network that can capture the artistic style of a diversity of paintings and demonstrated that such a network generalizes across the diversity of artistic styles.

- *Arbitrary styles per model* – research conducted by Huang et al. in [75] addressed limitations of Neural Style Algorithm introduced by Gatys et al. [63]. In particular, Huang et al. criticized the slow iterative optimization process, which directly influences practical applications of the Neural Style Algorithm which Gatys et al. proposed. The attempt to speed-up the style transfer process with a feed-forward network caused another issue – an inability to adapt to arbitrary new styles. An effective approach which enabled arbitrary style transfer in real-time was proposed by Huang et al. and at the heart of their method there was a novel adaptive instance normalization layer, which helped their model achieve speed comparable to the fastest existing approach. In essence, the model learns how to extract and apply any style to an image, which was the limitation of both single and a multi-layer style transfer models which were able to produce only images of those styles that they saw and learned during the training time.

All these findings assisted in our definition of the most suitable approach, considering both the limitations underlined in the recent publications, as well as the most advanced models developed so far. Cases for style transfer vary depending on the impact required from one specific industry, i.e., commercial art, gaming, virtual reality or photo editing applications. In our case, improving the existing photo editing tools by employing the style transfer was an initial inspiration, since they can be made very small and fast enough to run directly on mobile devices, which multiplies the range of potential applications.

Particularly relevant to our research is the interaction between high-level and low-level features. By changing the style of an image, or in other words modifying its low-level features, our goal is to preserve the high-level content. Thus, choosing the most appropriate images for the stylization process is a very important task and what makes an abstract art convenient for our research is the fact that it relies mostly on texture and color combinations. This is suitable when we want to target the automatic modification of low-level features, considering that:

- *Low-level features* are the texture, regions of the image, its edges, surfaces;
- *High-level features* include objects, scenes and events in an image.

While previous studies on automatic prediction of memorability from images paved the way towards the automatic recognition of image memorability, many questions are still open. For instance, is it possible to increase the memorability of an image while keeping its high-level content?

The link between memorability and style-related cues, like colors, has been previously explored and those studies reported that harmonious colors appear to be more memorable [76]. Other works offered a path towards an algorithmic understanding of not only the way humans create but also the way they perceive artistic imagery, by using the art theory of color combination to analyze emotions in abstract paintings [24].

Advances in computer vision research have not only allowed us to measure image memorability but have also demonstrated automatic prediction of this property through advanced deep learning models, providing the accuracy approximate to human performance [12]. These pioneering steps in the research and development of methods for analyzing image memorability have opened up the possibility of further combining data, i.e., style of photography, such as colors and textures, with its basic content in order to create models according to which memorability could be calculated, predicted and then automatically increased.

The pioneering work on Neural Style Transfer, described in [77], has been followed by many other studies, mostly addressing its limitations in terms of computational cost. In particular, study in [72] managed to reduce dramatically the time required for stylization, despite introducing the constraint that only a prefixed number of styles can be adopted. Some other researchers [73], proposed a style transfer method which is fast and works with arbitrary styles. Other research studies proposed modifications on the original style transfer framework [77] in order to adapt it to different applications, such as photorealistic [74] and semantic [25] style transfer.

More recent works [75] addressed the problem of improving the quality of the stylized images, considering second order statistics for style representation. Other works emphasized the problem of style transfer in videos [78], proposing methods which generate multiple frames with a specific style while ensuring temporal coherence. Common problem of applying a trained style transfer model to frames of a video is so-called flickering. Flicker is a visible change in brightness between cycles displayed on video displayers, caused by even the smallest changes and noise from frame to frame in a video. To solve this challenge, Gao et al. proposed a stable style transfer model in [79] able to generate coherent style transfer videos while maintaining the styles.

Some other works proposed an efficient technique for zero-short style transfer, like in [80], transferring arbitrary style into content images. Researchers proposed a style decorator which makes up the content features by semantically aligned style features from an arbitrary style image. They also started from a dilemma of the efficiency of style transfer for arbitrary styles and after conducting experiments, their results demonstrated the superiority of the method in generating arbitrary stylized images.

The complexity of learning models which have been developed to resolve different tasks kept rising as more and more fascinating models were created. The Generative Adversarial Network (GAN) belong to a group of models of this type, due to its ability to generate new data that closely resembles the examples used during training. This contributed largely to solving the challenging problems in computer vision, such as image-to-image translation and style transfer tasks. Since GANs are not only used by computer scientists but also by the artists, when used for artistic purposes, they are also called Creative Adversarial Networks (CANs).

Existing approaches using feed-forward generative networks for style transfer, either multi-style or an arbitrary-style transfer, usually suffer from poor image quality. Zhang et al. presented in [75] a Multi-style Generative Network (MSG-Net), which achieved both a real-time performance and a superior image quality compared to state-of-the-art approaches. The purpose of MSG-Net is to retain the functionality of earlier optimization-based approaches, while ensuring real-time processing and both goals were achieved.

Nevertheless, to the best of our knowledge, no previous works on deep stylization were focusing on modifying images in order to enhance specific high-level attributes, such as memorability.

Chapter 3

METHODOLOGY

In this section we introduce the proposed framework designed to automatically increase the memorability of an input image. We designed our method so that the process of generating highly memorable stylized images is performed in an efficient way. In its essence, the proposed approach exploits Neural Style Transfer methods to create stylized images, while preserving most of their high-level content. In our framework the process of stylization is driven by a specific module which ensures that the generated images have increased memorability and that most of the high-level content of the original images is preserved.

This section provides the methodological overview starting from the basic concepts that led to the development of our model. Then, the procedure used to train and test such a model is described, as well as the systematic and experimental activity conducted with the aim of obtaining a comprehensive understanding of how the model works. Finally, we present the role of datasets and training procedures in the final results, along with the details related to the size and the construction of all sets of data used during the research work on this thesis.

3.1. Overview

This dissertation focuses on exploiting knowledge from the most recent studies on memorability, from the collection of image datasets specifically designed to study memorability, to user-friendly techniques used to annotate these data with memorability scores. All this with an aim of setting up the framework for developing tools to automatically increase the memorability of photographs by changing and modifying their style by manipulating low-level features while preserving high-level representations.

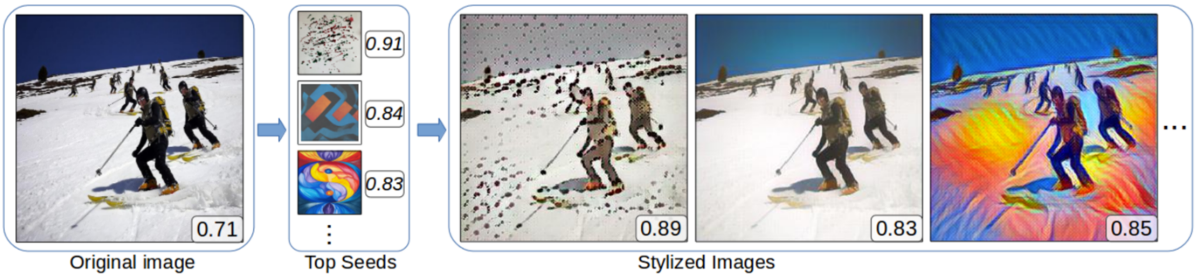


Figure 3.1: *Illustration of the main idea behind the proposed framework*

Figure 3.1. shows our proposed novel approach for increasing the memorability of images, inspired by editing-by-filtering framework. Given a generic, natural image (left – original image), our approach automatically finds the best style filters, i.e., the "style seeds" (central part of the image) which increase an overall memorability score of the input image the most. Style filters are sorted by corresponding predicted memorability increases. Memorability scores in the range [0,1] are placed in the bottom right corner of each image.

Our method relies on three deep networks. First one is the ***Synthesizer*** network, used to synthesize a memorable image from the input picture and a style picture. A second network acts as a style ***Selector*** and it is used to retrieve the "best" style seed to provide to the Synthesizer, (i.e., the one that will produce the highest increase in terms of memorability) given the input picture. To train the Selector, pairs of images and vectors of memorability gap scores (indicating the in- crease/decrease in memorability when applying each seed to the image) are used. A third network, the ***Scorer***, which predicts the memorability score from a given input image, is used to compute the memorability gaps necessary to train the Selector.

The starting point and the inspiration for the development of a new methodological framework and a special tool which will allow the automation of the entire process came from the recent discoveries from the Massachusetts Institute of Technology (MIT). A group of researchers from the Artificial Intelligence Lab have developed a method to modify the memorability of human face photographs [66] while maintaining the identity of persons and other facial features such as age, attractiveness and mapped emotional intensity, so-called emotional magnitude, [81] referring to the intensity of emotion expressed in faces.

The approach developed within research conducted for this thesis relies on the Neural Style Transfer method to create stylized images, with certain modifications which are necessary for enabling the memorability increase, along with the original image content preservation.

These two were the efficiency criteria used. Our approach co-articulates three main components, namely:

1) *The seed Selector (Sl, marked as* **R***)*
2) *The Scorer (Sc, marked as* **M***)*
3) *The Synthesizer (Sy -* **S***)*

Thus, we refer to it as ***S³*** or ***S-cube***. In order to give a general idea of the overall methodological framework. We illustrate the pipeline of ***S-cube*** in Figure 3.2.

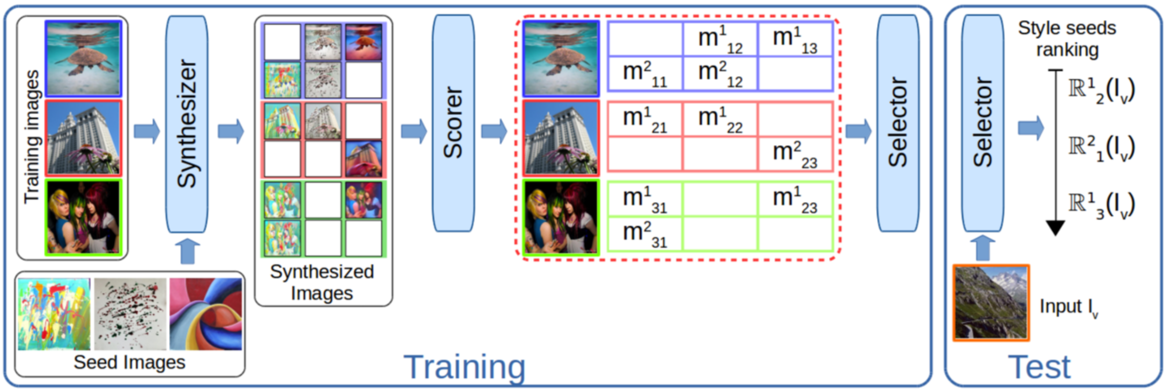


Figure 3.2: *An overview of method developed for the purpose of this research.*

At training time, the Synthesizer **S** and the Scorer **M** serve to generate the training data for the seed Selector **R**. Precisely, **S** and **M** are used to produce images from many input image-seed pairs and to score these pairs, respectively. Each input image is then associated with the relative increase or decrease of memorability for a new image and therefore rank the seeds according to the expected increase. The training data is highlighted with a red dotted frame in the Fig. 3.2. At test time, the seed Selector is able to retrieve a sorted list of style seeds for each new image, based on the predicted memorability increase $\mathbf{R}_{S}$ $(\mathbf{I}_{V})$, giving them further to the Synthesizer.

The Selector **R** is the core of the suggested approach: for a generic input image **I** and given a set of style image seeds *S*, the Selector retrieves the subset of *S* which will be able to produce the largest increase of memorability. In details, the seed Selector predicts the expected increase or decrease of memorability that each seed **S** ∈ *S* will produce in the input image **I** and consequently it ranks the seeds according to the expected increase of memorability. At training time, the Synthesizer and the Scorer are used to generate images from many input image-seed pairs and to score these pairs, respectively.

Each input image is then associated with the relative increase or decrease of memorability score obtained with each of the seeds. With this information, we can learn to predict the increase or the decrease of memorability for a new image, and therefore rank the seeds according to the expected increase. Indeed, at test time, the Selector is able to retrieve the most memoralizable seeds and give them to the Synthesizer.

We subsequently created an extended version of the original method where we learn how to predict not only the most memoralizable seeds but for each seed we also automatically compute the optimal degree of stylization.

3.2. The S-cube approach

Training Phase

During the training phase the three models are learned, the Scorer (*Sc* - **M**), the Synthesizer (*Sy* - **S**) and the seed Selector (*Sl* - **R**). The Scoring model *Sc* returns the memorability value *Sc*(**I**) of a generic image **I** and it is learned by means of training set of images annotated with memorability:

$$\mathcal{M} = \{I_i^M, m_i\}_{i=1}^I$$

In addition to this training set, we also consider a generating set of natural images:

$$\mathcal{G} = \{I_g^G\}_{g=1}^G$$

and a set of style seed images:

$$\mathcal{S} = \{S_s\}_{s=1}^S$$

The Synthesizer produces an image from an image-seed pair,

$$I_{gs} = Sy\left(\boldsymbol{I_g^G}, \boldsymbol{S_s}\right) \tag{1}$$

The Scoring model *Sc* and the Synthesizer *Sy* are the required steps to train the seed Selector *Sl*. Indeed, for each $I_g^G \in \mathcal{G}$ image and for each style seed $S_s \in \mathcal{S}$ the synthesis procedure generates I_{gs}.

The Scoring model is used to compute the memorability score gap between the synthesized and the original images:

$$m_{gs}^{Sc} = Sc(\boldsymbol{I}_{\boldsymbol{gs}}) - Sc(\boldsymbol{I}_{\boldsymbol{g}}^{\boldsymbol{G}}) \quad (2)$$

The seed-wise concatenation of these scores, denoted by $m_g^{Sc} = \left(m_{gs}^{Sc}\right)_{s=1}^{S}$ is used to learn the seed Selector. Specifically, a training set of natural images labeled with the seed-wise concatenation of memorability gaps is constructed:

$$\mathcal{R} = \{I_g^G, m_g^{Sc}\}_{g=1}^{G}$$

The process of seed selection is casted as a regression problem and the mapping **R** between the image and the associated vector of memorability gap scores is learned. Once the Selector is trained on **R**, it is able to estimate the vector of memorability gaps for a test image, which is much faster than running the Synthesizer and the Scorer *S* times (one per seed).

The memorability gap vector provides a ranking of the seeds in terms of their ability to memorabilize images (i.e., the best seed corresponds to the largest memorability increase).

Test Phase

During the test phase and given a novel image $\mathbf{I}_v$, the seed Selector is applied to predict the vector of memorability gap scores associated to all style seeds, i.e., $\mathbf{m}_v = Sl\ (\mathbf{I}_v)$. A ranking of seeds is then derived from the vector $\mathbf{m}_v$. Based on this ranking the Synthesizer is applied to the test image $\mathbf{I}_v$ considering only the top Q style seeds $\mathbf{S}_s$ produces a set of stylized images:

$$\{I_{qs}\}_{q=1}^{Q}.$$

3.2.1. The Scorer

The Scoring model *Sc* returns an estimate of the memorability associated to an input image **I**. In our work, we use the memorability predictor based on LaMem dataset in [12] which is the state of the art to automatically compute image memorability. LaMem is the largest annotated image memorability dataset to date, containing images from diverse sources with memorability scores from human observers. This dataset is almost 30 times larger than the previous one introduced by Isola et al. in 2014 in [39].

In details, following the same research, we consider the AlexNet CNN model as pre-trained in [82] named Hybrid-CNN. *AlexNet*[4] is the name of a Convolutional Neural Network (CNN) designed by Alex Krizhevsky et al. who competed in the ImageNet Large Scale Visual Recognition Challenge in 2012 and demonstrated that a deep convolutional network can be used for solving image classification problem. Today it represents one of the most influential papers in the field of computer vision (according to Google Scholar, as of 2020, AlexNet paper has been cited over 70.000 times). *Hybrid-CNN* was fine-tuned and trained to classify more than a thousand categories of objects and scenes.

Therefore, the network is pre-trained first for the object classification task (i.e., on ImageNet database) and then for the scene classification task (i.e., on Places dataset).

Then, we randomly split the LaMem training set into two disjoint subsets of 22,500 images each, $\mathcal{M}$ and $\boldsymbol{E}$.

We use the pre-trained model and the two subsets to learn two independent scoring models *Sc* and *Ev*. Starting from the pre-trained model, we minimize the Euclidean distance of the scores on each subset of LaMem, thus learning two independent scoring models *Sc* and *Ev*. While *Sc* is used during the training phase of our approach, the model *Ev* is adopted for evaluation, described in more details in the next Chapter. For training, we run 70k iterations of stochastic gradient descent with momentum 0.9, learning rate 10^{-3} and batch size 256.

3.2.2. The Synthesizer

The Synthesizer takes as input a generic image $\mathbf{I}_g$ and a style seed image $\mathbf{S}_s$ and produces a stylized image $I_{gs} = Sy(\mathbf{I}_g, Sy_s)$. In this work we consider two different neural style transfer methods to implement the Synthesizer, namely the approach in [83] and the most recent method in [73].

Ulyanov et al. emphasized the problem related to slow and memory-consuming optimization process presented by Gatys et al. [63] and proposed an alternative approach: given a single example of a texture, this approach trains compact feed-forward convolutional networks to generate multiple samples of the same texture of arbitrary size, as well as to transfer artistic style from a given image to any other image. In practical terms, Ulyanov et al. managed to maintain the quality of the textures, only generated hundreds of times faster.

[4] https://en.wikipedia.org/wiki/AlexNet

The strategy proposed by Ulyanov et al. in [83] consists of training a different feed-forward network for every seed. In this work, as seeds we use 100 abstract paintings from the DeviantArt database [84] and therefore train $S = 100$ networks for 10k iterations with learning rate 10^{-2}. The most important hyperparameter of the style transfer method in [83] is the coefficient α, which regulates the trade-off between preserving the original image content and generating something closer to the style image (see Fig. 3.3). In our experiments we evaluated the effect of α on the creation of highly memorable images, which we will describe in more details in the next Chapter 4.

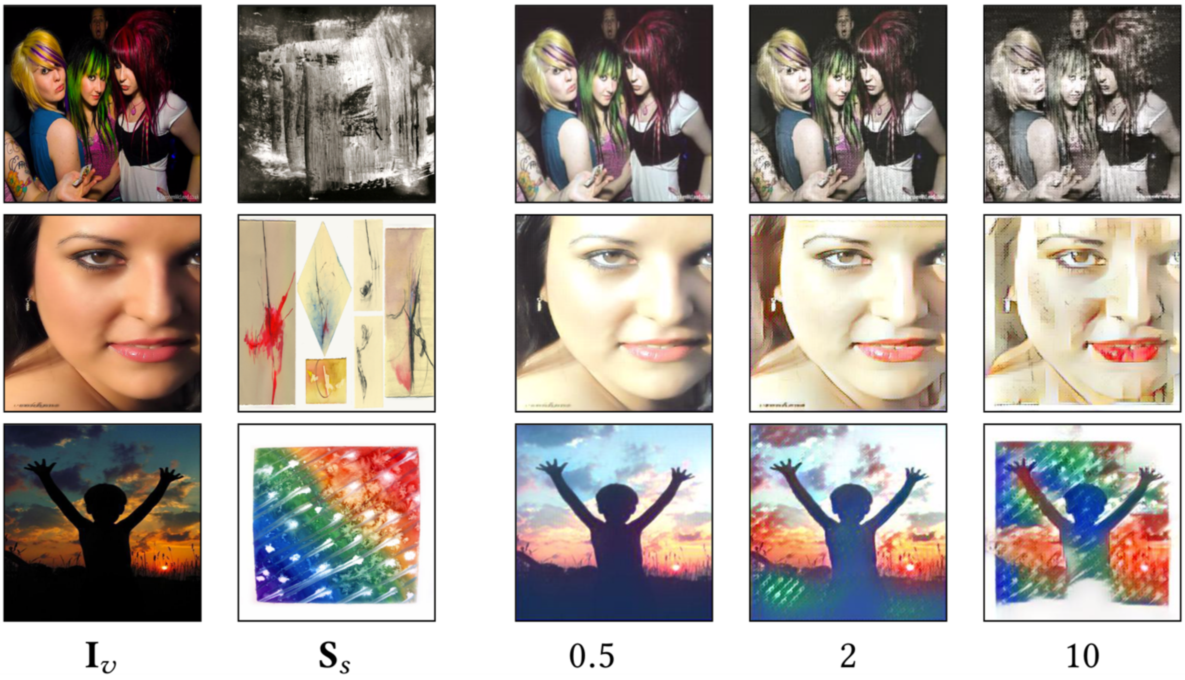


Figure 3.3: *Increasing image memorability with Neural Style Transfer: Sample stylized images*

In Figure 3.3., we present sample stylized images on the right, while the original images and applied style seeds are shown on the left side. Synthesized images are obtained with the method proposed by Ulyanov et al. in [83] at varying α.

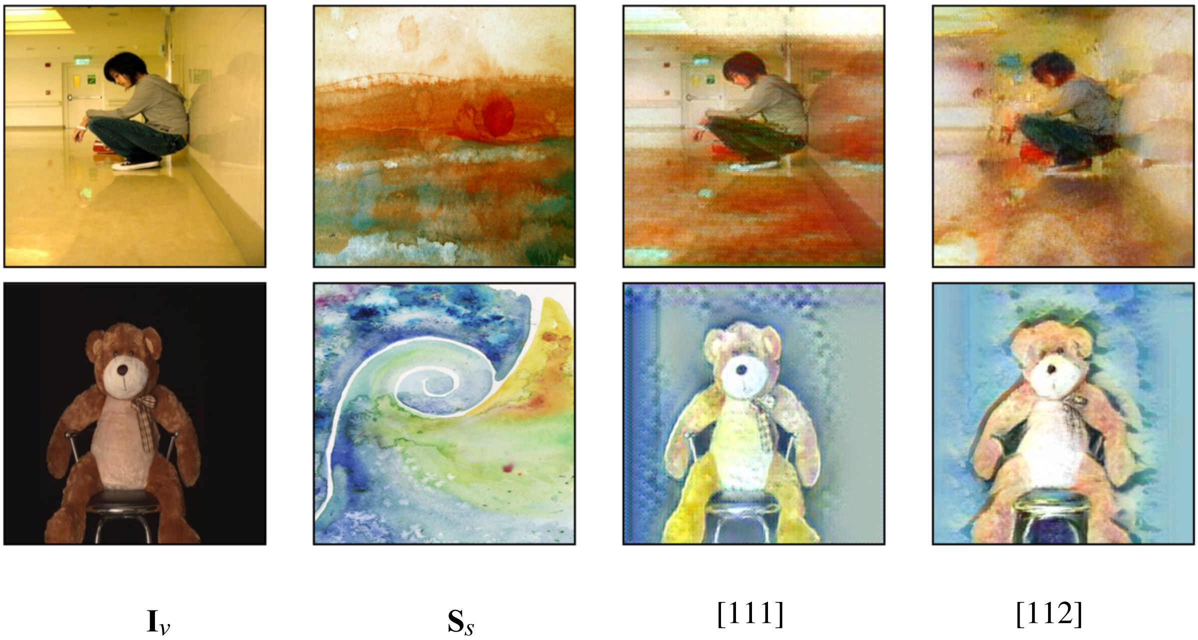


Figure 3.4: *Increasing image memorability by using Ulyanov et al. [111] and Huang et al. [112] methods.*

Figure 3.4. represents a sample stylized images, where original images are marked as $\mathbf{I}_v$, style seeds we applied are labeled as $\mathbf{S}_s$ and stylized images obtained using as our Synthesizer the methods described by Ulyanov et al. [83] and Huang et al. [73].

It is worth noticing that the methodology proposed in this research is independent of the synthesis procedure. Indeed, while working on the research tasks, we also tried some other methods, namely Gatys et al. [63] and Li et al. [79] as well recently proposed by Huang et al. [73], which achieves very good stylization performance while keeping the computational complexity low.

Method proposed by Gatys et al. in [63] focuses on neural image representations that are contributing to the semantic image understanding, or in other words understanding the meaning of the objects and the image scenery. And while a human brain is very good at achieving these results and understanding, there is a very complicated computation behind. Everything needs to be translated into a pixel information, which then needs to be transformed into patterns of neural activity.

The great success which managed to unlock the huge number of possibilities happened in 2012, where we had an artificial neural system able to perform the same computation as the human visual system, or at least an amazing performance that we had not seen before.

The system was introduced by Alex Krizhevsky et al. [80]. Its basic assumption, followed by a number of evidences from neuroscience, was that this great near-human performance happens thanks to the correspondence of the neural representations in the brain and the hierarchical structure of the single units which are performing this computation. In practical terms, this opened up a number of Image Manipulation possibilities, since it enabled our access to all the neurons and the complete neural representation of an input image. Consequently, we are able to change and modify the neural representation of the image.

This discovery opened up a question of finding a corresponding image that would lead to certain neural representation and by finding this image, we can achieve the desired change in the artificial neural network. CNN Gatys et al. used in the study is VGG 19-layer network, trained for object recognition in images. The network consists of two basic operations, one of which is called rectified convolution, with a small 3x3 convolution kernel and the second which is the downsampling mechanism based on Max-pooling.

Practically, what is happening between the input image and the first convolutional layer is that the neural network has a collection of features that are extracted from the input image. The amount of information about the actual pixel values that is preserved in feature representation could be visualized in feature maps and by trying to find a new image, the criteria Gatys et al. decided to set was to find the same feature maps. In the lower layers, this information about the actual pixel values of the input image is preserved (pictures correspond to the original one), but by observing the higher layers, it is easy to spot the information loss (about the actual pixel values), while the object observed in the image, even in the higher layers, is still there, as well as the general structure of the scenery. Essentially, they use these features in the higher layers to extract the content of the image.

The second part of the algorithm is a texture model (related to the style of the image) that Gatys et al. build on top of the feature maps. After preserving the content, the second task of the neural style transfer is to extract the texture (i.e., the colors, the structure of the image). Gatys et al. do this by computing correlations between the feature maps, in each of the layers of the image.

Another method we considered for this research was proposed by Huang et al. [73], previously described as capable of achieving a very good stylization performance while keeping the computational complexity low, which is important in our framework since the Synthesizer is also used to generate the training set for learning *Sl*.

Images are depicting some sample stylization results obtained using the two style transfer methods considered in work [83] and [73].

3.2.3. The Seed Selector

Given a training set of natural images labeled with the vector of memorability gaps:

$$\mathcal{R} = \{\boldsymbol{I}_g^G, \boldsymbol{m}_g^{Sc}\}_{g=1}^G$$

the seed Selector *Sl* is trained minimizing the following objective:

$$\mathcal{L}_{Sl} = \sum_{g=1}^{G} \mathcal{L}\left(Sl(\boldsymbol{I}_g^G), \boldsymbol{m}_g^{Sc}\right) \qquad (3)$$

In this equation, the $\mathcal{L}$ is a loss function which measures the discrepancy between the learned vector $Sl(\boldsymbol{I}_g^G)$ and memorability gap scores m_g^{Sc}.

This approach has several advantages:

- *First*, we can very easily rank the seeds by the expected increase in memorability they will produce if used together with the input image and the synthesis procedure.
- *Second*, if several seeds have similar expected memorability increase, they can be proposed to the user for further selection.
- *Third*, if all seeds are expected to decrease the memorability, the optimal choice of not modifying the image can easily be made.
- *Fourth*, once *Sl* is trained, all this information comes at the price of evaluating *Sl* for a new image, which is cheaper than running *Sy* and *Sc S* times.

Although this strategy has several advantages at testing time, the most prominent drawback is that, to create the training set **R**, one should ideally call the synthesis procedure for all possible image-seed pairs. This clearly reduces the scalability and the flexibility of the proposed approach. The scalability because training the model on a large image dataset means generating a much larger dataset (i.e., *S* times larger). The flexibility because if one wishes to add a new seed to the set **S**, then all image-seed pairs for the new seed need to be synthesized and this takes time. Therefore, it would be desirable to find a way to overcome these limitations while keeping the advantages described in the previous paragraph.

The solution to these issues comes with a model able to learn from a partially synthesized set, in which not all image-seed pairs are generated and scored. This means that the memorability gap vector m_g^M has missing entries. In this way we only require to generate enough image-seed pairs. To this aim, we propose to use a decomposable loss function $\mathcal{L}$. Formally, we define a binary variable ω_{gs} set to 1 if the gs-th image-seed pair is available and to 0 otherwise and rewrite the objective function in (3) as:

$$\mathcal{L}_{Sl} = \sum_{g=1}^{G} \sum_{s=1}^{S} \omega_{gs\mathrm{l}}\left(Sl_s\left(\boldsymbol{I}_g^G\right), \boldsymbol{m}_g^{Sc}\right) \tag{4}$$

where Sl_s is the s-th component of *Sl* and 1 is the square loss. We implement this model using an AlexNet architecture, where the prediction errors for the missing entries of $\boldsymbol{m}_g^{Sc}$ are not back-propagated. After exploring approaches of combining predictions of numerous models with an aim of reducing errors and after considering the cost in time-consuming terms, we discovered an efficient version of model combination, presented as a technique called "dropout" which consists of setting to zero the output of each hidden neuron with probability 0.5. The neurons which are "dropped out" do not contribute to the forward pass and do not participate in back-propagation. The reason Krizhevsky et al. [80] employed this regularization method was inspired with an idea of reducing overfitting in the fully-connected layers while training a DNN to classify the 1.2 million high-resolution images.

To do so in our research, we considered the model presented in [82] by Bolei et al. where they pre-trained Hybrid CNN by combining the training set of Places CNN and ImageNet CNN, removing the overlapping scene categories and provided a visual representation of the CNN layers' responses (Fig. 3.5), demonstrating that an object-centric network using ImageNet and a scene-centric network using Places, learn different features. This helped us understand differences in the internal representations of object and scene networks.

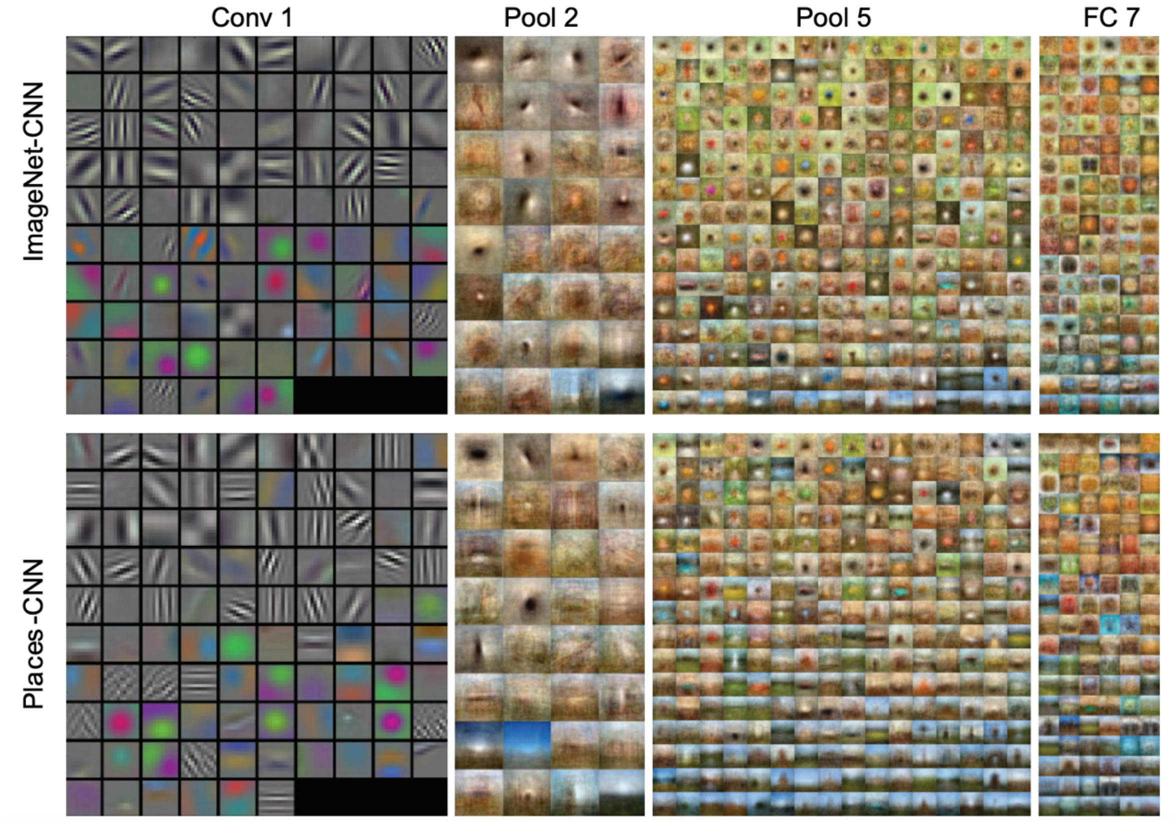


Fig. 3.5: *Visualization of the units' receptive fields at different layers for the ImageNet CNN and Places CNN presented by Bolei et al. in* [82]. *Conv 1 units contain 96 filters. The Pool 2 feature map is 13x13x256. The Pool 5 feature map is 6x6x256. The FC7 feature map is 4096x1.*

Acknowledging that the deep features from the response of the Fully Connected Layer - fc7 of the CNNs, the final fully connected layer before producing the class prediction, have only minor difference between the feature of the fc6, we decided to fine-tune only the layers fc6, fc7, conv5, conv4 using a learning rate equal to 10^{-3}, momentum equal to 0.9 and batch size 64. Our choice of Hybrid-CNN is considered more appropriate when dealing with generic images in our case, due to the fact that the network is pre-trained both on images of places and objects.

3.2.4. Learning the Degree of Stylization

In the ***S-cube*** method described above, the parameter α used by the Synthesizer and regulating the trade-off between content and style is assumed to be fixed and is defined a priori. In this section we introduce an extended version of ***S-cube***, where the seed Selector is also able to predict the optimal α value, i.e., the optimal degree of stylization, for a given image-seed pair.

To this aim we modify the implementation of the seed Selector and train it by using an augmented training set generated with the Synthesizer.

Specifically, by considering explicitly α we rewrite the previous equation in (1) denoting the stylized image as $I_{gs}^{\alpha} = Sy\left(I_g^G, S_s, \alpha\right)$. The memorability score gap between the synthesized and the original image in (2) is denoted as $m_{gs\alpha}^{Sc} = Sc\left(I_{gs}^{\alpha}\right) - Sc\left(I_g^G\right)$ and $m_{g\alpha}^{Sc}$ indicates the vector of the score gaps for all styles and a given value of α.

Assuming that the set $\mathcal{A}$ of the possible values of α is finite, for implementing the Selector we propose to learn different deep network models, each corresponding to a specific degree of stylization α and computing for each image $\mathbf{I}_g$ the associated memorability gap vector $\boldsymbol{m}_{g\alpha}^{Sc}$.

However, if the cardinality of the set $\mathcal{A}$ is large, learning different models may be inefficient both in terms of memory and computational cost. To address this issue, we resort to a multi-task learning framework, inspect the pretrained Hybrid CNN model and implement the different network models, restricting them to share the same parameters in the convolutional block and differ only in the fully connected layers. Our assumption is that the learned models are related as their training sets have been generated with the same style seeds and images.

In our experiments we considered three different values of α, i.e., $\alpha \in \mathcal{A} = \{0.5, 2, 10\}$.

This means that, for instance, for generating the augmented training set for the Selector with the method in [83] we use $S = 300$ networks, one for each pair of α value and style $\boldsymbol{S}_s$.

3.3. CNN Architecture

Fully connected neural networks or dense neural networks consist of multiple hidden layers and each of the hidden layers is densely connected to its previous layer. Densely connected precisely means that every input is connected to every output in that layer (Fig 3.6).

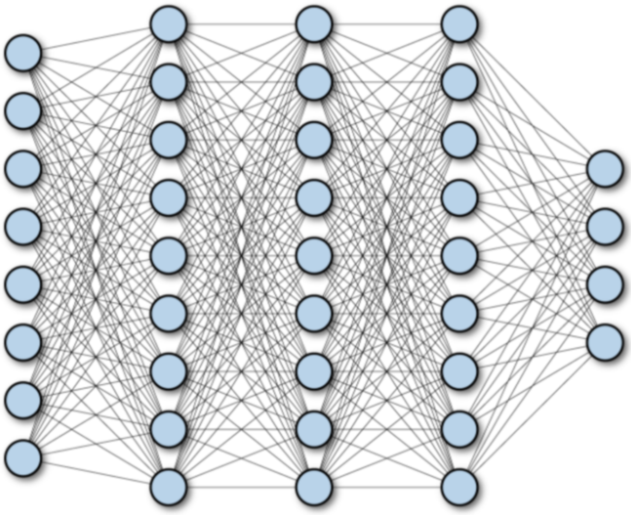


Figure 3.6: *Fully Connected Neural Network*

Using these networks for the image classification tasks essentially means that we are taking the two-dimensional spatial structure, the input image, and translating into one-dimensional vector, which we can then feed through the CNN, making sure that every pixel in that vector will feed into the next layer. Practically, we are connecting every single pixel of that input image to every single neuron in our hidden layers. Due to this complexity when it comes to practical implementations, we are trying to build spatial structure, images, into neural networks in a more rational way, with an aim of facilitating our learning process. Considering at the same time the importance this spatial structure has in image data, our goal is to maintain this structure and we are doing this by connecting patches of the input to a single neuron in the hidden layer. So, instead of connecting every input pixel with, from an input image to a single neuron in a hidden-layer, we are connecting just a small patch and this way we can achieve our goal – to learn visual features by simply waving those connections in the patches. Instead of connecting these patches uniformly to the CNN hidden layer, we are going to weight each of the pixels, or more commonly we use the weighted summation of all those pixels in that patch, which feeds into the next hidden unit in our hidden layer, to detect a particular feature. In practice, this is called *convolution*, the first main part of CNN.

Apart from extracting the features from our input images, or from the previous layers, there are the other two important parts of CNNs. The second part is applying the *Non-linearity*, helping us to deal with nonlinear data and also to introduce complexity into our learning pipeline for solving even more complex tasks. The third step is *Pooling operation*, which allows us to down sample our image spatial resolution and deal with multiple scales of the features of that image.

The presented architecture can be used for many other applications, considering that the second part of the pipeline, the feature extraction part, can be used to attach any kind of output that we want.

Therefore, this class of artificial neural networks became dominant in various computer vision tasks, designed to automatically learn these spatial hierarchies of features. However, its performance can be influenced with the selection of the datasets. Small datasets which were previously used for basic recognition tasks were not suitable for our research, considering our inspiration and the initial idea to develop a model suitable for arbitrary images, which have many more variables than those captured in small datasets.

The development of AlexNet has been an inspiration to many researchers, our team included. We decided to implement our model by using an AlexNet architecture, pre-trained for two different tasks on two different sets of data, composed of hundreds of thousands to millions of labeled images.

Essentially, the architecture of AlexNet consists of eight layers, out of which (Fig. 3.6):

- five convolutional layers
- three fully-connected layers.

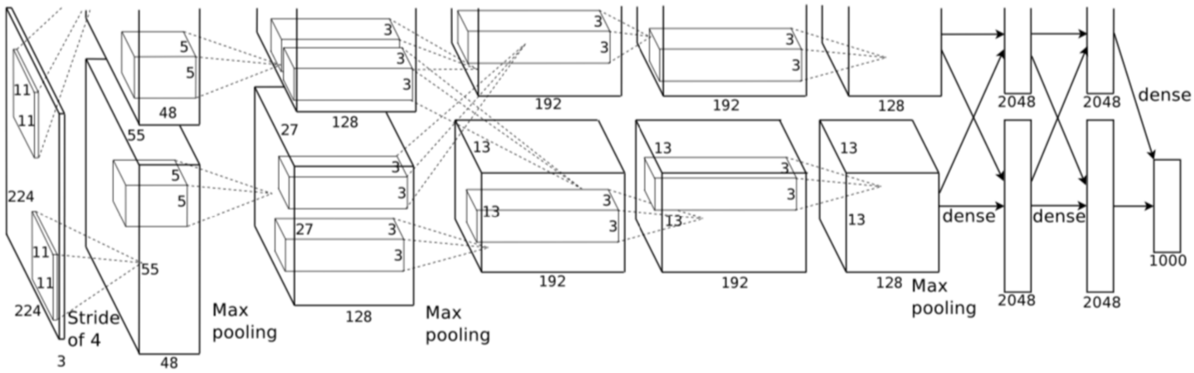


Fig. 3.6: *Illustration of AlexNet architecture presented in* [80]

The reason why convolutional neural networks became the neural networks of choice for many data scientists, partly is influenced by the hierarchical feature extraction ability but more importantly by their huge improvement in performance. The choice is also driven by the ability to tune different parameters. In the last few years, CNNs have been successfully applied to identify faces, objects and traffic signs as well as powering vision in robots and self-driving cars.

Nevertheless, numerous researchers use so-called hybrids of the already proposed architectures, with an aim of improving deep CNN performance. In case of our research the choice of Hybrid-CNN [82] is considered more appropriate when dealing with random images, thanks to the fact that the network was pre-trained both on images of places and objects.

The second architecture used for the purpose of our research is VGG16 (Fig. 3.7), used for tasks of classification and detection; this network model was originally proposed by researchers from University of Oxford. It has been reported that this model achieves 92.7% accuracy in ImageNet and that it makes improvement over AlexNet, which is why we pre-trained VGG16 on ImageNet and AlexNet (Hybrid-CNN).

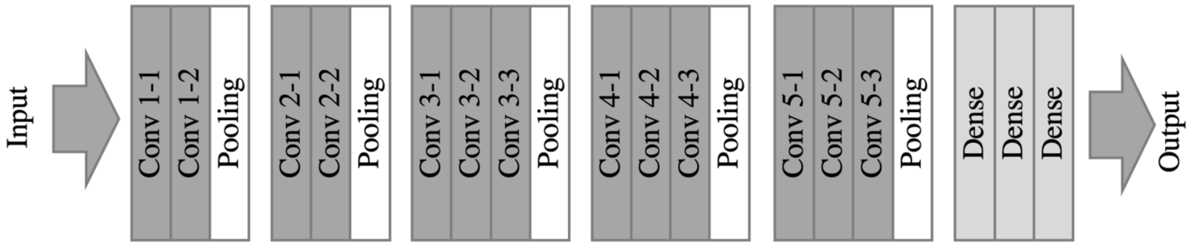


Fig. 3.7: *VGG – 16 Layer Definition*

As presented in Fig. 3.7, there are 13 convolutional layers, 5 Max Pooling layers and 3 Dense layers which sums up to 21 layers but only 16 weight layers.

In order to explore more precisely a property like image memorability, a number of tasks has to be performed in order to build an image memorability model, starting from understanding the intrinsic-extrinsic factors and the fact that some images are more memorable than others, over its context and observers' behavior. Finally, the effectiveness of the novel solution has to be estimated and evaluated. Designing a system which will translate image-related data into some useful insights requires different sets of data which we will present in the upcoming lines.

3.3.1. ImageNet Dataset[5]

The ImageNet project started in 2009 with an aim of creating an easily accessible image database which will serve and benefit the entire visual object recognition research community. It was inspired by a growing tendency in the image and vision research field, the need for more data. Indeed, this image database consists of more than 14 million human-annotated images of each concept, divided in 22.000 visual categories. The photographs were annotated by humans using crowdsourcing platforms such as Amazon’s Mechanical Turk.

Since 2010, ImageNet project also runs an annual contest, or a challenge called The ImageNet Large Scale Visual Recognition Challenge (ILSVRC), whose purpose is to evaluate algorithms for object detection and image classification tasks at large scale. The sets of data used for the challenge contain approximately 1 million images and 1.000 object classes.
The purpose is to allow researchers to:

- Compare progress in detection across a wider variety of objects, at the same time taking advantage of the expensive labeling effort.
- Measure the progress of computer vision for a large-scale image indexing for retrieval and annotation.
- Participate in a workshop and computer vision conferences, where the methods and the results of the challenge are presented.

They are typically doing this by performing tasks:

1) Image classification: predicting the classes of objects present in an image.
2) Single-object localization: on top of image classification task, drawing a bounding box around one example of each object present.
3) Object detection: image classification and a requirement to draw a bounding box around each object present.

In 2012, AlexNet CNN achieved top-5 error of 15.3% more than 10.8% points lower than the runner up (26.1%). What facilitated this incredible achievement was the use of the GPUs, Graphic Processing Units during the training time.

[5] http://www.image-net.org/

3.3.2. Places Dataset[6]

In order to reach human-level performance with machine learning algorithms, we normally face two challenges:

- the algorithm must be suitable for the task i.e., CNNs in Large Scale Visual Recognition;
- an access to a training dataset.

Places set of data was designed inspired and guided by human visual cognition principles, with an aim of building a core of visual knowledge which could be used to train artificial systems for high-level visual understanding tasks, such as scene context, object recognition, action and event prediction. It consists of over 10 million scene photographs comprising over 400 unique scene categories, provided for academic research and education purposes, together with the trained CNNs. The Places dataset features 5000 to 30.000 training images per class, consistent with real-world frequencies of occurrence.

The dataset has different categories of bedroom, or streets, since they belong to different semantic categories which are defined by their function. The different function of these images consequently changes the way to make predictions, as it is hard to know for sure what can happen next in each of the places illustrated in these pictures. Therefore, the Places dataset is labeled with 434 scene semantic categories, comprising about 98% of the type of places a human can encounter in the world.

Using convolutional neural networks (CNNs), Places dataset provides scene classification CNNs (Places CNNs) as baselines, which significantly outperforms previous approaches. The aim of developing this dataset was to enable and promote learning of deep scene features for various scene recognition tasks, primarily used in academic research and education purposes.

3.4. Data and Setup

The efficiency of the proposed research, that is, of a new model developed for automatic memorability enhancement, was evaluated through experiments on a publicly available

[6] http://places2.csail.mit.edu/

database of approximately sixty thousand images, more precisely 58.741 images (LaMem). In addition, considering our research motivation of keeping the high-level content out of the modification process, the DeviantArt database, composed of 500 abstract paintings has also been taken into account. The reason for the latter choice is practical, due to the fact that such images do not contain much information, other than the texture and color combination, which makes them ideal for the automatic process of modifying low-level features.

3.4.1. LaMem Dataset[7]

Large Scale Image Memorability abbreviated LaMem represents the largest annotated image memorability dataset to date. LaMem contains nearly 60.000 images from diverse sources and it was built by the group of researchers at MIT Massachusetts Institute of Technology, Khosla et al. [12] with an aim of making progress in estimating visual memorability, which has been limited by the small scale and lack of variety of benchmark data. To overcome this, a novel experimental procedure has been introduced in order to progress the most objective measures of human memory.

Using Convolutional Neural Networks (CNNs), Khosla et al. show that fine-tuned deep features outperform all other features by a large margin, reaching a rank correlation of 0.64, near human consistency (0.68). Responses of the high-level CNN layers are demonstrating the exact objects and regions which are positively and negatively correlated with memorability, which allowed Khosla et al. to create memorability maps for each image and provide a method to perform memorability of images from various classes.

After conducting visual memory game in their research paper [12] where human participants were observing images (shown on the screen one after another) and signalizing each time when they detect the image that has already been presented (the repeating image), Khosla et al. were able to define memorability as the proportion of times in which a repeat of one image on the screen is correctly detected. To simplify this memorability measure, if an image was shown to 100 people out of which 80 of them correctly detected the repeat, then the memorability score will be 80/100=0.8. The impressive thing discovered that even if the experiment is repeated over and over again, the memorability score remains consistent, which means that memorability is property of an image, rather than the individual.

[7] http://memorability.csail.mit.edu/demo.html

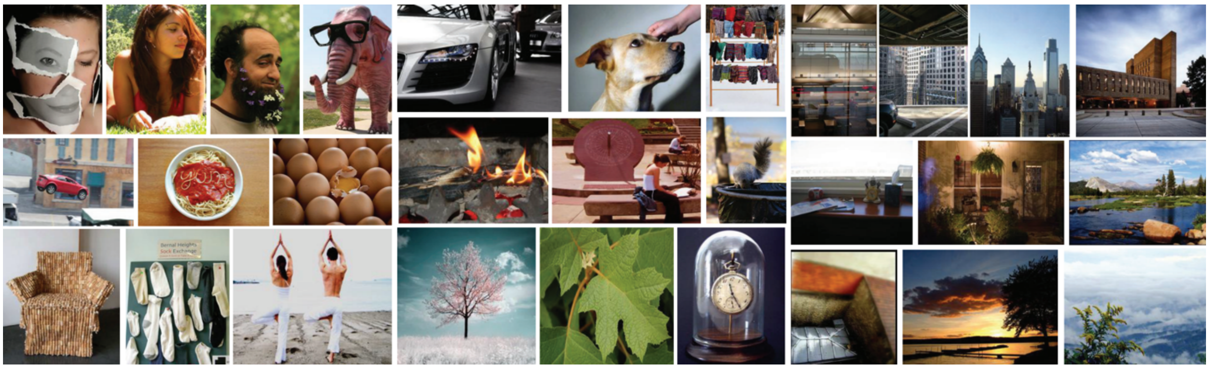


Figure 3.8*: Sample images from LaMem arranged by their memorability score (decreasing from left to right). LaMem contains a very large variety of images ranging from object-centric to scene-centric images, and objects from unconventional viewpoints. Images from* [12]

For predicting memorability, Khosla et al. had a set of training images with associated memorability scores. Their goal here was to build the function (f), that when given a test image, predicts how memorable it is. In this phase the CNN for Memorability was used with a typical architecture structured with layers. The great part of this process is that it requires only the output to be specified and it learns what the representation should be, without humans providing supervision. Compared with the human performance of 0.68 rank correlation, this memorability network obtained a rank correlation of 0.64, thus, very high to the human-level performance score.

To help us understand what really happened in this process, Khosla et al. visualized neurons and it was clear the way they feed the network with the number of images, while recording the activations at different parts of the network. Once all the activations were recorded, it is easy to extract the image which managed to activate some particular neuron. For instance, a neuron in Conv 4 that one image activated to a large extent. Still, it is hard to say which particular part of the image caused the activation. What is required at this stage is a batch to be taken from the image, a noise batch, which is then translated across the image. Based on this translation, it is easy to see how the activation is affected for that particular neuron and consequently, to understand which part of the image is important. All this can later help us find the receptive fields of different neurons across different layers and then project that back to images, in order to spot the discriminative parts of the images.

Essentially, what has been found is that the early layers learn very simple things, such as colors and simple edges (Fig.3.9).

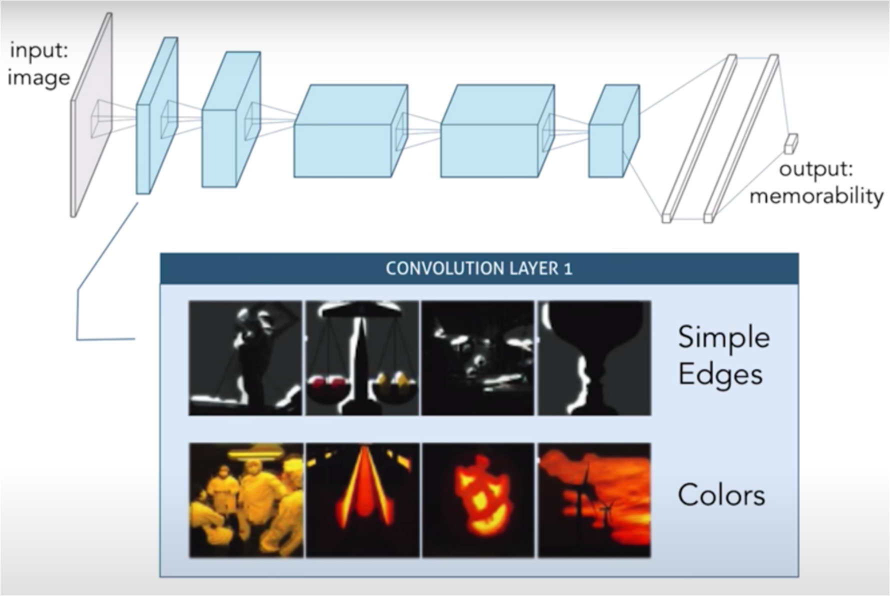


Figure 3.9: *Convolutional Layer 1 learning simple edges and colors*

The next layer learns complex edges and textures. Going further through the network, we can spot the shapes, object parts, and at the last ones even the scenes and full objects.

By visualizing neurons, it is possible to see the positive and negative neuron correlation with memorability, or in other words, to visualize also the images and understand their memorable representations. Strongly positive correlation Khosla et al. found in images of body parts, human faces and objects, while the most forgettable images were the ones containing the landscapes.

After obtaining these results and defining the tool that is able to predict memorability, it is simple to understand the memorability of any image and how it is measured, by applying this tool to a random image batch by batch and obtaining scores for the image. On the LaMem Demo website, Khosla et al. presented this as a heat map to demonstrate which parts of a randomly used image are more or less memorable (Fig. 3.10 and Fig. 3.11).

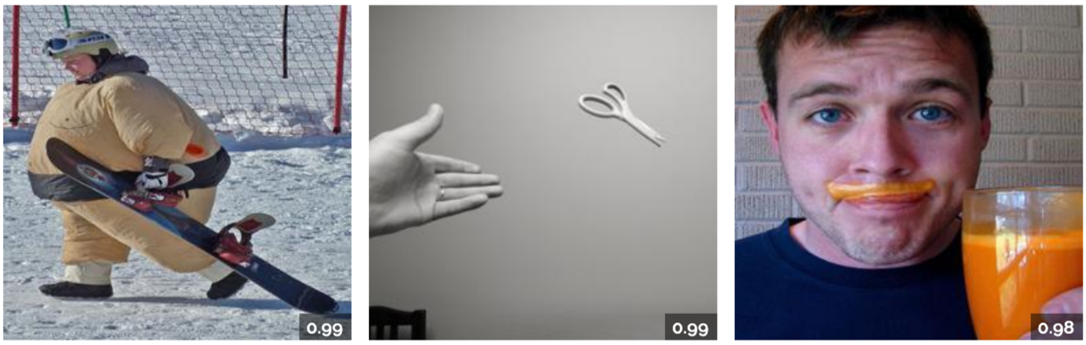


Figure 3.10: *LaMem Demo with memorability scores*

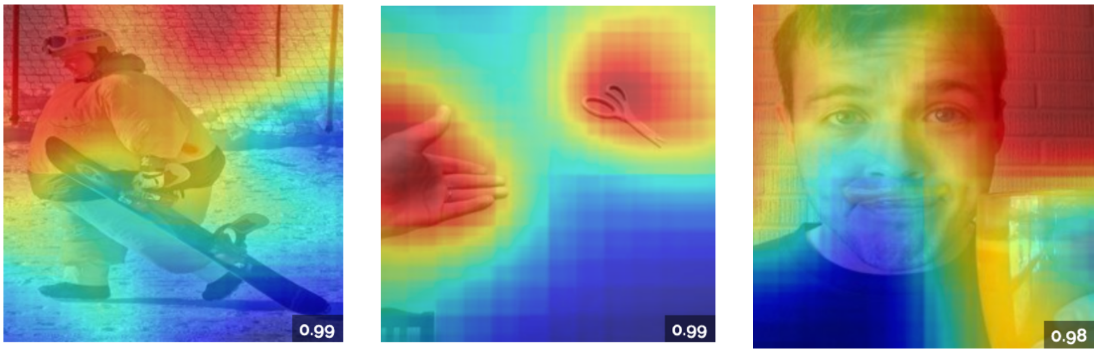


Figure 3.11: *LaMem Demo with heat maps showing the most memorable parts (red)*

LaMem is a collection gathered from a number of previously existing datasets, including the Affective Images dataset [85] which consists of Art and Abstract paintings. The International Affective Picture System (IAPS), widely used in emotion research; consists of the natural color photos depicting complex scenes containing different objects and scenes which typically provoke emotional reactions i.e., attack scenes, insects, accidents, pollution, babies and more.

Understanding how images can affect people on emotional level and how emotions also arise in the viewer of an image, Machajdik et al. expressed their curiosity in [85] and an idea of enabling image retrieval based on their emotional content and consequently, they developed methods for the low-level feature extraction and combination which represent the emotional content of an image. The Affective Images dataset creation was inspired by the theoretical and empirical concepts from both Psychology and Art Theory. The potential of applications Machajdik et al. are mentioning in their research is interesting, i.e., magazines searching for illustrative photos for articles covering different emotional or psychological concepts and so demonstrating the “depressed person” or just an image with a “sad” atmosphere.

3.4.2. DeviantArt Dataset

DeviantArt (dA) set is composed of 500 abstract paintings collected from an online social network site devoted to sharing, dissemination and marketing of user-generated artworks. It represents one of the largest online art communities with around 360 million artworks and over 35 million registered users, both amateur and professional artists. All artworks are organized in categories like photography, digital art, traditional art, et cetera, which simplifies the download of one specific genre only. For the purpose of our study, we decided to collect the artworks of abstract art, from the category Traditional – Art – Paintings – Abstract Art.

In the case of DeviantArt images selection, the main intention was to try to avoid substantial modifications of the high-level image content, thus, we preferred using the style seeds from abstract paintings, linked and associated with the texture and color combinations exclusively, and therefore, specifically appropriate for the automatic modifications of low-level features.

3.4.3. Dataset Construction

From the two datasets, considered the most suitable for the purpose of this research, we used a total set of 45.000 images which were then randomly split in two disjoint subsets of 22.500 images each.

The scores of an overall image memorability were collected for all the pictures in the dataset, using an optimized protocol of the memorability game.

Analyzing the responses of the high-level CNN layers, we were also able to understand which objects and regions are positively and negatively correlated with memorability, which provided a basis on which to perform further image memorability manipulation. Details are available in the Chapter 4 dedicated to the experimental evaluation.

3.5. Memory Game and User Study

In addition to quantitative evaluation, we also conducted qualitative evaluation, by creating a memory game. The inspiration for this comes from the previously published papers as well as from our interest to test and perhaps improve proficiency in cognitive and memory skills, but more importantly it served as an additional validation.

In our research, we followed the visual memory game protocol presented by Khosla et al. in [12] which consists of a stream of images presented on the screen, one after another, each for 1 second. The task of the users is to press the button each time they see an image that has been presented before during the experiment. Based on these experimental results, they defined a metric called memorability. Khosla et al. used Mechanical Turk workers to play their Memory Game.

The participants in our memory game were students reached through the University offices and various student associations, in total 200 students. Children were excluded from the evaluation process. Further details are available in Chapter 4.

Chapter 4

EXPERIMENTAL VALIDATION

Here we report the role and the performance of ***S-cube*** in increasing image memorability by selecting the most memoralizable seed styles. Later on, we introduce the results of the user study we conducted. Through it we managed to collect the actual memorability scores for a set of original and stylized image pairs. Lastly, qualitative results are presented.

4.1. Experimental setup

This research focuses on the analysis of more datasets, namely LaMem with nearly 60,000 images and DeviantArt set of 500 abstract paintings, with the idea of overcoming previous limitations in works related to memorability that were using small sets of data.

During the experimental phase, when the LaMem dataset is concerned, we used **45000 images for training, 10000 images for testing** and **3741 images for validation**.

While doing this, we split the LaMem training set into two subsets of 22,500 images each, $\mathcal{M}$ and $\boldsymbol{E}$, which were used to train two predictors *Sc* and *Ev*, respectively. The model *Sc* is the Scorer in our framework, while *Ev* (which we will denote in the following as the external predictor or the evaluator) is used to evaluate the performance of our approach, as a proxy for human assessment. We highlight that *Sc* and *Ev* can be used as two independent memorability scoring functions, since $\mathcal{M}$ and $\boldsymbol{E}$ are disjoint.

The validation set is used to implement early stopping. To evaluate the performance of our Scorer models *Sc* and *Ev*, following Khosla, et al. [12], we compute the rank correlation between predicted and actual memorability on the LaMem test set. We obtain a rank correlation of 0.63 with both models, while [12] achieves a rank correlation of 0.64 training on the whole LaMem training set. As reported in [12], this is close to human performance (0.68).

The test set of LaMem, composed of 10k images, is then used to learn the proposed seed Selector and to evaluate the overall framework and the Selector in particular. In detail, we split the LaMem test set into train, validation and test for our Selector with proportion 8:1:1, meaning 8,000 for training and 1,000 for validation and test. The training set for the Selector was already introduced as $\mathcal{G}$. The validation set is used to perform early stopping, if required. We denote the test set as $\mathcal{V}$.

Regarding the seeds, we estimated the memorability of all paintings of DeviantArt using *Sc* and selected the 50 most and the 50 least memorable images as seeds in the set $\boldsymbol{S}$. The memorability scores of the DeviantArt images range from 0.556 to 0.938.

For the user study, we randomly selected 66 images from $\mathcal{V}$ and we assigned to each of them a style image randomly extracted from $\boldsymbol{S}$. Then, for each image-seed pair we collected the actual memorability scores following the memorability game protocol described in [12].

The purpose of our user study is twofold:

- to show that our method is able to increase the memorability of arbitrary images and
- to demonstrate that the external predictor *Ev* is a good proxy for a user evaluation.

4.2. Evaluation metrics

We evaluate the performance of our method in predicting the memorability increase of an image-seed pair using two different performance measures:

- the mean squared error *MSE*, defined as:

$$MSE^X = \frac{1}{SV}\sum_{s=1}^{S}\sum_{v=1}^{V}\left(m_{vs}^{X} - Sl_s(\boldsymbol{I}_v^{\boldsymbol{V}})\right)^2 \qquad (5)$$

- and the accuracy A, defined as:

$$A^X = \frac{1}{SV}\sum_{s=1}^{S}\sum_{v=1}^{V}(1 \; - |HS\;(m_{vs}^X) - HS(Sl_s(\boldsymbol{I}_v^V))|) \qquad (6)$$

where the generic image I_v^V is taken from a set of (yet) unseen images $\mathcal{V} = \{I_v^V\}_{v=1}^V$ and the seed *S* is taken from a set of style seeds. *HS* is the Heaviside step function, which sets the input variables to 0 or 1, respectively when their initial values are lower or higher than zero [86]. The evaluation is performed based on the internal predictor *Sc*, the external predictor *Ev* or the human assessment *H,* as indicated with X ∈ {*Sc, Ev, H*}.

4.3. Image Manipulations Baseline

For all we know, this is the first in-depth study which achieves an automatic increase of memorability of an arbitrary chosen image, which is exactly why a directly explicit and quantitative analogy with past studies is not feasible. Evidently, what recent works succeeded at [12] was exactly what paved the way for the future image manipulations – computing accurately memorability maps from images.

Starting point was to use different sets of individual features in order to establish one such map, later substituted with a weighted pooling combination in order to deliver an overall memorability map for each individual image [87].

Furthermore, a memorability map was used also for removing details in an image, concretely through a cartoonization process. This is a popular artistic form which commonly contains an artistic abstraction. However, after the implementation of these processes, the outputs were typically resulting with a decrease of an overall memorability. Oppositely, the purpose of our work is to effectively increase memorability of a picture without modifying its high-level content, which is why previously described approach is not comparable with ours.

The only existing procedure which has a potential of being compared to our approach is [66], with an exception that the model was constructed specifically for face photographs, thus, its principle can't be directly transferred to a generic image. As a result, we construct an average baseline $\mathfrak{B}$ based on ranking the style seeds according to the average memorability increase for the training set, defined as:

$$\bar{m}_s^{Sc} = \frac{1}{G}\sum_{g=1}^{G} m_{gs}^{Sc} \tag{7}$$

Particularly, we are comparing the proposed image-dependent seed selector with an image-independent seed selector. The latter consists in selecting, for a test image, the seed that maximizes the memorability gap on average over the generating set.

	A^{Sc}		A^{Ev}		MSE^{Sc}		MSE^{Ev}	
$\underline{\omega}$	$\mathcal{B}$	***S-cube***	$\mathcal{B}$	***S-cube***	$\mathcal{B}$	***S-cube***	$\mathcal{B}$	***S-cube***
0.01	**63.21**	57.12	**60.96**	56.01	**0.0113**	0.0138	**0.0119**	0.0137
0.1	64.49	**64.70**	61.07	**62.22**	**0.0112**	0.0114	**0.0117**	0.0119
0.5	64.41	**67.18**	61.06	**64.38**	0.0112	**0.0102**	0.0117	**0.0106**
1	64.41	**67.75**	61.06	**64.70**	0.0112	**0.0102**	0.0117	**0.0108**

Table 4.1: ***S-cube*** compared with baseline $\mathcal{B}$ at varying percentage of training data $\underline{\omega}$

We compared the performance of ***S-cube*** with the baseline at varying percentage of training data, measured in terms of Accuracy A demonstrated in the left side of the Table 4.1 and Mean Square Error (MSE), presented in the right half of the same table. The achievements have been evaluated using both the internal *Sc* and the external *Ev* predictor.

	A^{Ev}		MSE^{Ev}	
S	$\mathcal{B}$	***S-cube***	$\mathcal{B}$	***S-cube***
20	60.66	**63.15**	0.0114	**0.0111**
50	61.09	**63.61**	0.0116	**0.0109**
100	61.06	**64.38**	0.0117	**0.0106**

Table 4.2: ***S-cube*** compared with baseline $\mathcal{B}$ at varying the cardinality *S* of the style set $\boldsymbol{\mathcal{S}}$ measured in terms of Accuracy A^{Ev} and Mean Square Error MSE^{Ev}.

4.4. Experimental Results

In this section we examine the effectiveness of our method and its capability and potency in suggesting “memoralizable” style seeds under different experimental setups.

For training individual components of the proposed ***S-cube*** approach, we employed different libraries to build and train neural networks, while relying also on diverse deep learning frameworks. In such a way, Synthesizer *Sy* is based on Pytorch, a Python-based scientific computing package. On the other hand, the Scorer *Sc* is based on Convolutional Architecture for Fast Feature Embedding known as *Caffe*[8], originally developed at University of California, Berkeley. It is an open source deep learning framework written in C++ with a Python interface. As for the training procedure, we started from previously mentioned MemNet, with an exception of splitting training part of the dataset in two parts, in order to train two independent Scorers. Finally, our seed Selector *Sl* is based on Theano and Lasagne Python libraries. The complete procedure along with our code for reproducing the results are publicly available on the development platform GitHub.com.

4.4.1. Increasing Image Memorability

Table 4.1 displays the performance of the proposed approach (***S-cube***) and the baseline ($\mathfrak{B}$) for different values of the average amount of image-seed pairs $\underline{\omega}$. Specifically, $\underline{\omega} = 1$ means that all image-seed paris are used, $\underline{\omega} = 0.1$ signalizes that only 10% is used, and so forth.

The stated accuracy (A) and the MSE were evaluated using the internal scoring model *Sc* and the external scoring model *Ev*. Broadly, the method that we propose outperforms the baseline in case there is enough image-seed pairs, as explained earlier, deep architectures require a sufficient amount of data in order to be effective. Surely, when $\underline{\omega} = 0.01$, the network optimization procedure attempts to learn a regression model from the raw image to a 100-dimensional space with, on average, only one of these dimensions propagating the error back to the network. In other words, although this dimension varies for each image, it’s different for every single photograph we use, it still might happen that not enough information is propagated

[8] Caffe model is available on the link: https://yadi.sk/d/059Y2cii3SQ39L

back so as to effectively learn a robust regression. This situation is coherent when the scoring method changes from *Sc* to *Ev*. An increase in performance is evident when using *Sc*.

Since the seed Selector has been trained to learn the memorability gap from *Sc*, the performance is higher when using *Sc* instead of *Ev*. This result, further, motivates the need of having an external Scorer *Ev*, trained on an independent set of images, to evaluate the performance of our method. In this series of experiments and in the following, unless otherwise specified, we set $\alpha = 2$.

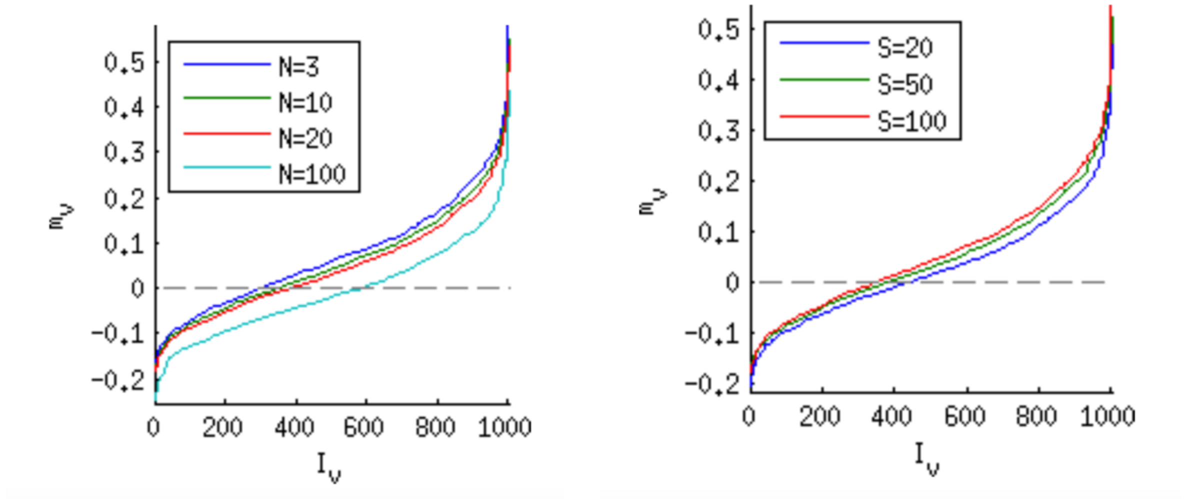


Fig 4.1: *Sorted average memorability gaps* $\underline{m}_v$

These sorted average memorability gaps $\underline{m}_v$ are obtained with method ***S-cube*** shown on the left side of the graph (Fig. 4.1.), averaging over a varying number of top *N* seeds and in the right side of the graph over varying cardinality *S* of the seed set, with $N = 10$.

Abscissa corresponds to test image index, ranked by $\underline{m}_v$. The wider is $\underline{m}_v$, the better.

Moreover, we studied the performance and the behavior of our framework when varying the size *S* of the seed set. Results are shown in Table 4.2. The parameter $\underline{\omega}$ is set to 0.5. Precisely, we select two sets of 50 and 20 seeds out of the initial 100, randomly sampling these seeds half from the 50 most memorable and half from the 50 least memorable ones.

In terms of accuracy, the performance of both the proposed method and the baseline stays fairly stable when decreasing the number of seeds. Yet, a different trend is observed for the MSE. While the MSE of the proposed method increases when reducing the number of seeds (as expected), the opposite trend is found for the baseline method. We argue that, even if the baseline method is robust in terms of selecting the best seeds to a decrease of the number of seeds, it does not perform well at predicting the actual memorability increase. Instead, the proposed method is able to select the best seeds and measure their impact better, especially when more seeds are available. This is especially important if the method wants to be deployed with larger seed sets.

We also assess the validity of our method as a mechanism for increasing the memorability of a generic input image *Iv* in the most effective way.

In Figure 4.1 on the left graph we report the average memorability gap $\underline{m}_V$ over the top N seeds retrieved, with $N = 3, 10, 20$ and all the seeds.

For display purposes, we rank the images of test set $\mathcal{V}$ by their average memorability gap, so that the curve closer to the upper-left corner corresponds to the best method (Fig. 4.1).

It can be noted that $\underline{m}_V$ achieves higher values when smaller sets of top $\boldsymbol{N}$ seeds are considered, as an indication that our method effectively retrieves the most memorabilizing seeds.

In the same figure (Fig. 4.1.) in the right graph we report the average memorability gaps $\underline{m}_V$ obtained over the test set $\mathcal{V}$ with ***S-cube***, considering $N = 10$ and a varying number of style seeds S. As expected, a larger number of seeds results in a higher increase in terms of memorability.

Lastly, we investigated the link between the memorability score of the style seeds and the corresponding average memorability increase obtained for each seed. A low correlation is found between these two sets, meaning that the seed style alone is not predictive for the memorability increase and that the optimal ranking is dependent also on the image. In detail, we found a Pearson rank correlation $\varrho=0.277$ in the case of $\alpha = 2$. These results further motivate the intuition behind the proposed method: increasing the memorability of an image means selecting the style which better matches its visual appearance.

4.4.2. S-cube as a generic framework

In this subsection we demonstrate some of our additional results in order to show that the proposed ***S-cube*** is a general framework and that different choices can be made to implement the main system components. Precisely, in our experiments, we take into a consideration the Synthesizer and the Selector.

As previously mentioned, for the Synthesizer we use different style transfer methods, namely Ulyanov et al. [83] and Huang et al. [73]. Performance in terms of Accuracy and MSE for ***S-cube*** and the baseline $\mathcal{B}$ are reported in Table 4.3. With the respect to the baseline $\mathcal{B}$, better results are obtained by using our ***S-cube*** approach.

	A^{Ev}		MSE^{Ev}	
	$\mathcal{B}$	***S-cube***	$\mathcal{B}$	***S-cube***
Ulyanov et al. [83]	61.06	**64.38**	0.0117	**0.0106**
Huang et al. [73]	70.08	**75.12**	0.0142	**0.0112**
AlexNet [80]	61.06	**64.38**	0.0117	**0.0106**
VGG16 [88]	61.06	**64.39**	0.0117	**0.0111**

Table 4.3. Performance of ***S-cube*** compared to the baseline $\mathcal{B}$ under different implementation choices, evaluated in terms of Accuracy and MSE and using the external predictor (*Ev*)

Top scores presented in Table 4.3 are obtained by using the style transfer methods in Ulyanov et al. [83] and in Huang et al. [73]. Bottom scores, as written, by using different architectures for the Selector.

These results demonstrate that our method can integrate different style transfer methods, suggesting that ***S-cube*** can be easily upgraded (and further improved) when novel style transfer techniques are made available in literature.

Similarly, for the Selector we consider different deep networks. In particular, in Table 4.3 we report the performance of the ***S-cube*** when using two different architectures, VGG16 (pre-trained on ImageNet) and AlexNet (Hybrid-CNN). From the same table it can be observed that ***S-cube*** consistently outperforms the baseline.

Furthermore, we investigated the impact of using different style transfer techniques on the style ranking for each image in the test set $\mathcal{V}$. To this aim, we computed the Pearson correlation between the memorability gaps obtained with ***S-cube*** using different style transfer methods. These values are reported in Table 4.4. As it can be seen, a strong correlation, i.e., ϱ =0.875, is found when considering Huang et al. and Ulyanov et al. with α = 2. A correlation value greater than 0.8 is found when considering Ulyanov et al. with different α values.

	Huang et al.	Ulyanov et al.	Ulyanov et al.	Ulyanov et al.
		α=0.5	α=2	α=10
Huang et al.	1.000	0.789	0.875	0.938
Ulyanov et al. α=0.5	-	1.000	0.936	0.836
Ulyanov et al. α=2	-	-	1.000	0.934
Ulyanov et al. α=10	-	-	-	1.000

Table 4.4: Pearson correlations between memorability gaps of each image-seed pair in the test set $\mathcal{V}$, obtained using different style transfer methods.

4.4.3. Analysing the impact of the degree of stylization α

We also performed additional experiments to study the impact of the hyper-parameter α on the performance of the method. In these experiments we consider the style transfer approach in Ulyanov et al. [83] for implementing the Synthesizer.

In Table 4.5 we show the performance of ***S-cube*** when the system is trained using a predefined α value.

		A^{X}		MSE^{X}	
	α	$\mathcal{B}$	***S-cube***	$\mathcal{B}$	***S-cube***
X=*Sc*	0.5	62.70	**63.37**	0.0102	**0.0100**
	2	64.41	**67.75**	0.0112	**0.0102**
	10	67.99	**73.25**	0.0125	**0.0104**
X=*Ev*	0.5	58.30	**59.50**	0.0107	**0.0103**
	2	61.06	**64.70**	0.0117	**0.0108**
	10	68.31	**71.71**	0.0132	**0.0111**

Table 4.5. Performance of ***S-cube*** at varying α.

Performance of ***S-cube*** is reported in terms of Accuracy and MSE using both the internal predictor (*Sc*) and the external one (*Ev*).

$\mathcal{A}$	Top 3	Top 10	Top 20	Top 30	All
{0.5}	0.0574	0.0377	0.0217	0.0143	-0.0085
{2}	0.0739	0.0567	0.0440	0.0352	-0.0096
{10}	0.0695	0.0651	0.0573	0.0493	-0.0251
{0.5, 2, 10}	**0.0742**	**0.0688**	**0.0606**	**0.0516**	**0.0064**

Table 4.6. Average memorability increases obtained when considering the top N memoralizable style seeds retrieved with ***S-cube*** and different sets $\mathcal{A}$.

In all the cases ***S-cube*** performs better than the baseline $\mathcal{B}$, both when considering the internal and the external predictor.

In Table 4.6 we present the performance obtained with the extended version of our method which estimates the best style seed and the Synthesizer parameter α for a given image. We report the average memorability increases over the test set $\mathcal{V}$, obtained when averaging over the top N best style seeds retrieved for each test image ($N = 3, 10, 20, 30$ and 100).

The best performance of our method is achieved when it is possible to choose the optimal α value for each image-seed pair. In other words, by actively selecting α it is possible to achieve a higher increase in terms of memorability.

In order to give an idea of how well our method is performing, we computed the Upper Bound (UB) of the memorability increase, achievable for a given set of images and styles, in the case of α =2 and Top 3. Specifically, we run the Stylizer with for the set of image-style pairs, we ranked for each test image the obtained stylized images according to the memorability measured by the internal Scorer *Sc*, we selected the Top 3 for each test image and we averaged the corresponding memorability increases based on the external Scorer *Ev*. The following results are obtained: Baseline ($\mathcal{B}$): 0.0694; ***S-cube***: 0.0739; UB: 0.0804.

To further analyze the impact of the degree of stylization, we also conduct an experiment to see if there is correlation between high/low values of the parameter α and certain types of style seeds. More precisely, we run an experiment to verify whether it exists for each seed a tendency to be assigned to a specific α value among the three considered (α =0.5, 2 or 10). The probability for each style seed to be assigned to α equal to 0.5, 2 or 10 is depicted in Fig.4.2. It is easy to see that this probability changes according to the style seed. In other words, in order to increase the memorability of a test image, some seeds tend to be selected with a low α, others with a high value.

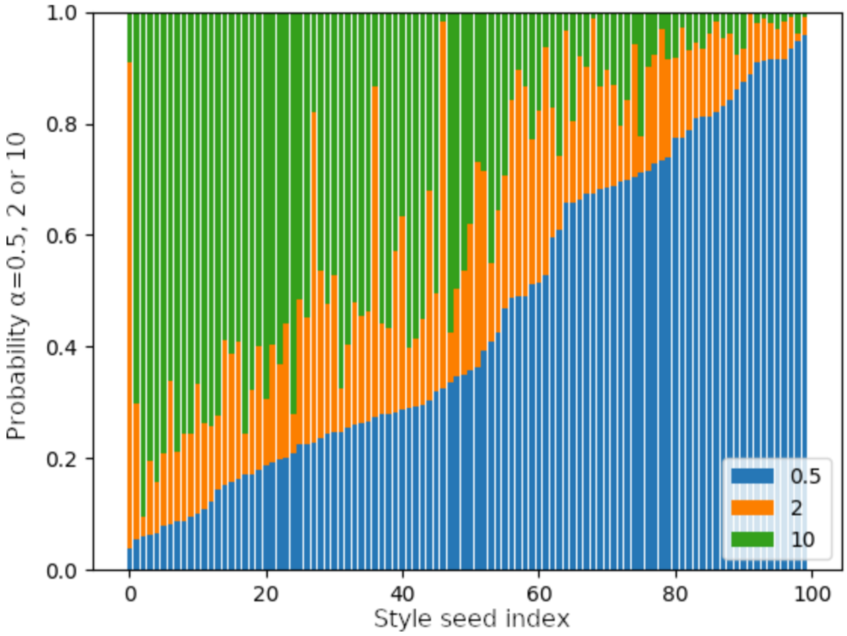


Fig 4.2: *Bar plot*

Our bar plot (Fig. 4.2) is showing, for each of the 100 style seeds used, the probability to be selected with α =0.5 in blue color, for α =2 in orange color and for α =10 by our method ***S-cube*** shown in green color. The seed styles are sorted by crescent probability of being selected with α =0.5.

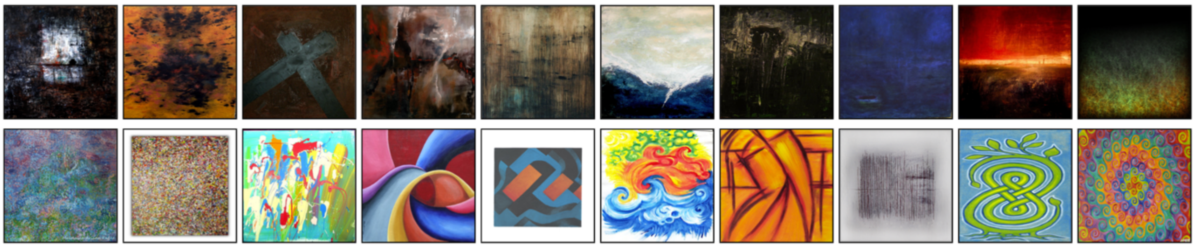


Fig. 4.3*: (Top) Top 10 seeds having a higher probability to be assigned to α =0.5; (Bottom) Top 10 seeds having a higher probability to be assigned to α =10.*

In Fig. 4.3, we reported the top 10 style seeds which are usually assigned with a low value of stylization coefficient (α = 0.5) (see Fig.4.2-top) and the top 10 styles which are usually assigned with a high value of stylization coefficient (α = 0.5) (see Fig.4.2-bottom).

Indeed, it can be seen that images from the first set look mostly dark and gloomy, while those from the second set are more colorful and bright. Also, in the first set, lines and edges are less defined with respect to the second set.

4.5. User study

Committed also to understanding the users' behavior and motivations, we decided to additionally investigate and demonstrate the effectiveness of our method by conducting a user study. In this section we report the results of the analysis of our feedback methodology by incorporating additional experimental and observational research methods. First, we provide the details of the memory game protocol that we implemented in order to collect memorability scores from users and then we illustrate the results of our study.

4.5.1. Memory game protocol

We collected the actual memorability scores for a verification set $\mathcal{Z}$ which was randomly sampled from the test set $\mathcal{V}$, as described in the earlier sections. To do this, we followed the protocol of the efficient memory game, described in [12], applying some modifications in order to adapt the game to our scenario of image stylization.

Our memory game session was built by randomly selecting 66 target images, 12 vigilance repeats and 59 fillers out of the 1,000 images of the LaMem test set $\mathcal{V}$. For each image, we randomly sampled an associated style out of $\mathcal{S}$. For each session, we made sure that 33 targets are shown in their original version, while the other 33 are displayed in their corresponding stylized version. Also, we make sure that each image is displayed strictly in its original or in its stylized version, with α=2. Each target repeat is shown after a minimum of 35 to a maximum of 150 images. Vigilance repeats were shown within 7 images from the first showing. Vigilance, in modern psychology, is defined as the ability to maintain concentrated attention over a prolonged period of time [89]. In our case, vigilance repeats were ensuring that the user is focused on the game. While for each session the set of targets, vigilance repeats and fillers is preserved, we created 100 image sequences with a different image sorting, randomly assigned, so that no memory bias is introduced by showing each image at different times of the game session. When a new player starts the game, a random sequence among the possible 100 is considered.

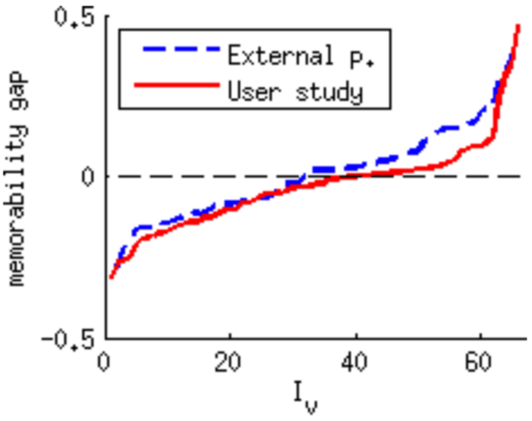


Fig. 4.4: *Sorted memorability gaps for the image-seed pairs in the set **Z** measured using the predicted and actual memorability scores collected through the user study.*

	A^X		MSE^X		ϱ^X	
	$\mathfrak{B}$	***S-cube***	$\mathfrak{B}$	***S-cube***	$\mathfrak{B}$	***S-cube***
X=*Sc*	**72.73**	65.15	0.0155	**0.0138**	0.517***[9]	**0.570*** **
X=*Ev*	**68.18**	63.64	0.0183	**0.0162**	0.479***	**0.544*** **
X=*H*	**60.61**	59.09	0.0188	**0.0153**	0.191*[10]	**0.453*** **

Table 4.7. *Performance of our method compared to baseline **B** evaluated on the set **Z** using the internal predictor Sc, the external predictor Ev and the actual memorability scores collected through the user study H. Performances are reported in terms of Accuracy (A), MSE and Pearson's correlation ρ.*

In order to recruit volunteers for the game, we sent invitation emails to several mailing lists of university groups and trusted associations, connected with both University of Novi Sad and University of Trento.

[9] ***p-value < 0.005

[10] *p-value > 0.05

Before accessing the game, participants were asked to provide personal information, such as age, gender and nationality (the latter was optional). We recruited over 200 players, both male and female (respectively 42.3% and 57.7%), aged between 17 and 62 years old, and from a large variety of countries. Some of the countries participants were coming from are: Austria, Belgium, Bosnia and Herzegovina, Brazil, Chile, Croatia, Estonia, France, Germany, Greece, Italy, India, Iran, Latvia, Lithuania, South Korea, Malaysia, Mexico, Mongolia, Montenegro, Netherlands, Poland, Portugal, Romania, Russia, Serbia, Singapore, Spain, Sweden, Turkey, U.K, Ukraine, U.S.A., and more.

Each volunteer played only once and all the results of those players who did not qualify to the game, by detecting less than 25% - i.e., less than 4 - of the vigilance repeats, were discarded from the further analysis.

Finally, for each of the 132 target images (66 original and the corresponding 66 stylized versions), we computed the actual memorability scores by aggregating the performance of over 80 players.

4.5.2. Results

The internal and external predictors are trained in natural images. One fair question is whether or not the outcome of these predictors is still valid for stylized images. In particular, we would like to assess the correlation between the external predictor (used to evaluate the proposed approach) and the human scores. To this aim we compute the values of the Pearson's correlation coefficient and MSE considering the automatically predicted and the actual memorability scores of the images in the verification set $\mathcal{Z}$. The Pearson's correlation values corresponding to the original and the stylized images in $\mathcal{Z}$ are respectively $\varrho = 0.66$ and $\varrho = 0.50$ (we recall that human performance is $\varrho = 0.68$). We also computed the Spearman's rank correlation, which systematically leads to the same conclusions as the Pearson's correlation coefficient.

These results demonstrate a high correlation between actual and predicted memorability scores. For the stylized set, a drop in performance can be observed which can be probably ascribed to the domain shift existing between original and stylized images. Indeed, the Scorer model is trained on the LaMem dataset which does not include stylized images. In fact, while LaMem includes abstract art images, they represent a small subset. Thus, the learned memorability predictor will be obviously more accurate in the case of generic images. A similar trend can be observed for the MSE: 0.0121 for the original images and 0.0132 for the stylized ones. Indeed, a 16.7% increase in the error is motivated by the lower accuracy of the external predictor on the stylized set.

Furthermore, in Figure 4.4 we plot the sorted memorability gaps obtained with the external predictor and the user study for the 66 image-seed pairs in the verification set. Overall, a similar trend can be observed, with the external predictor which only slightly overestimates the memorability gaps.

Finally, in Table 4.7 we show the performances of our method compared to the baseline $\mathcal{B}$ in predicting the memorability increase for the images in our verification set $\mathcal{Z}$. Note that in this case, differently from previous experiments, for each image in the verification set we only have a single corresponding style and therefore we consider only the associated predicted memorability score.

The performance is measured in terms of Accuracy, MSE and Pearson's ϱ, and evaluated using the internal predictor (*Sc*), the external predictor (*Ev*) and the actual memorability scores collected though the user study (*H*). It can be seen that ***S-cube*** outperforms the baseline in all cases in terms of MSE and ϱ.

In the case of Accuracy, an advantage is observed for the baseline. A possible interpretation of these results is that, for this particular set of images, $\mathcal{B}$ probably performs better in predicting the direction of the memorability increase, while it still performs worse at estimating the absolute value of the variation. This is shown by the low Pearson's coefficient ϱ, which reaches non statistically significant values in the case of *H*.

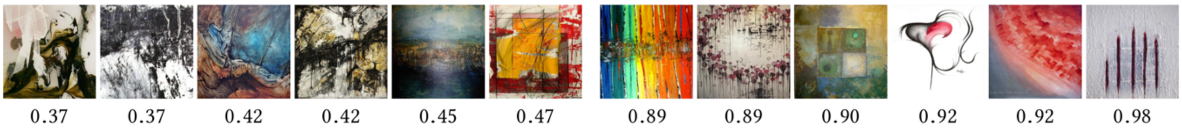


Fig. 4.5: *The most (right side) and the least (left side) memorable abstract art paintings from LaMem dataset with their actual memorability scores, reported below each image.*

4.6. Qualitative Analysis and Discussion

Our approach to understanding the phenomena relies also on the analysis of qualitative data. In this section, rather than predicting or explaining the data, we will emphasize future insights based on the link between memorability and style, more precisely, our focus will be on the link between memorability and complexity, which has been poorly explored up to now, compared to the link between memorability and some other attributes like interestingness or colors, as described in the related works chapter. Moreover, some visual results from the two sets of data, respectively the abstract art images and the data used in our user study, will be discussed.

4.6.1. Abstract art and memorability

In this section we will focus on the set of abstract art images included in the LaMem dataset, annotated with their actual memorability scores. The purpose of this is to analyze also qualitatively the specific patterns which can explain differences in memorability.

As explained in the earlier sections, this image set allows us to target only on the visual style of images, discarding high level semantics which naturally influence memorability, like in the case of face photographs or presence of other objects who tend to make images more memorable. In detail, we extracted from the LaMem dataset the 280 abstract art paintings which are part of the Affective dataset. Following those logics, we manually discarded from this set those paintings containing people or objects, in order to completely cut out any semantic information. After this process we selected a total of 187 images, with a memorability score ranging from 0.37 to 0.98.

In Figure 4.6 we show the five least and the five most memorable abstract art images, together with the corresponding memorability scores.

In the images on the left side (Fig. 4.6) we observe some common patterns, such as higher complexity, darker colors, absence of straight lines, while those on the right tend to be simpler and with more pleasant colors.

This finding suggests that complexity and dark colors may have a negative influence on memorability. In other words, it appears that simple paintings, especially those with pleasant colors, tend to be more memorable. This may be explained with the fact that complex patterns require more time to be assimilated by the viewer. In the memorability game, each image is displayed for exactly the same amount of time: complex patterns probably get a lower chance to be decoded and remembered. This finding is in line with previous studies in the literature [90].

In the next page we present some sample results of the user study (Fig. 4.6.) where for each block (left) original input image, (center) style seed and (right) corresponding synthesized image. The actual memorability scores and gaps are reported respectively below each original and stylized image. Both cases where style transfer produces an increase (L1-L8) and a decrease (R1-R8) in memorability are shown.

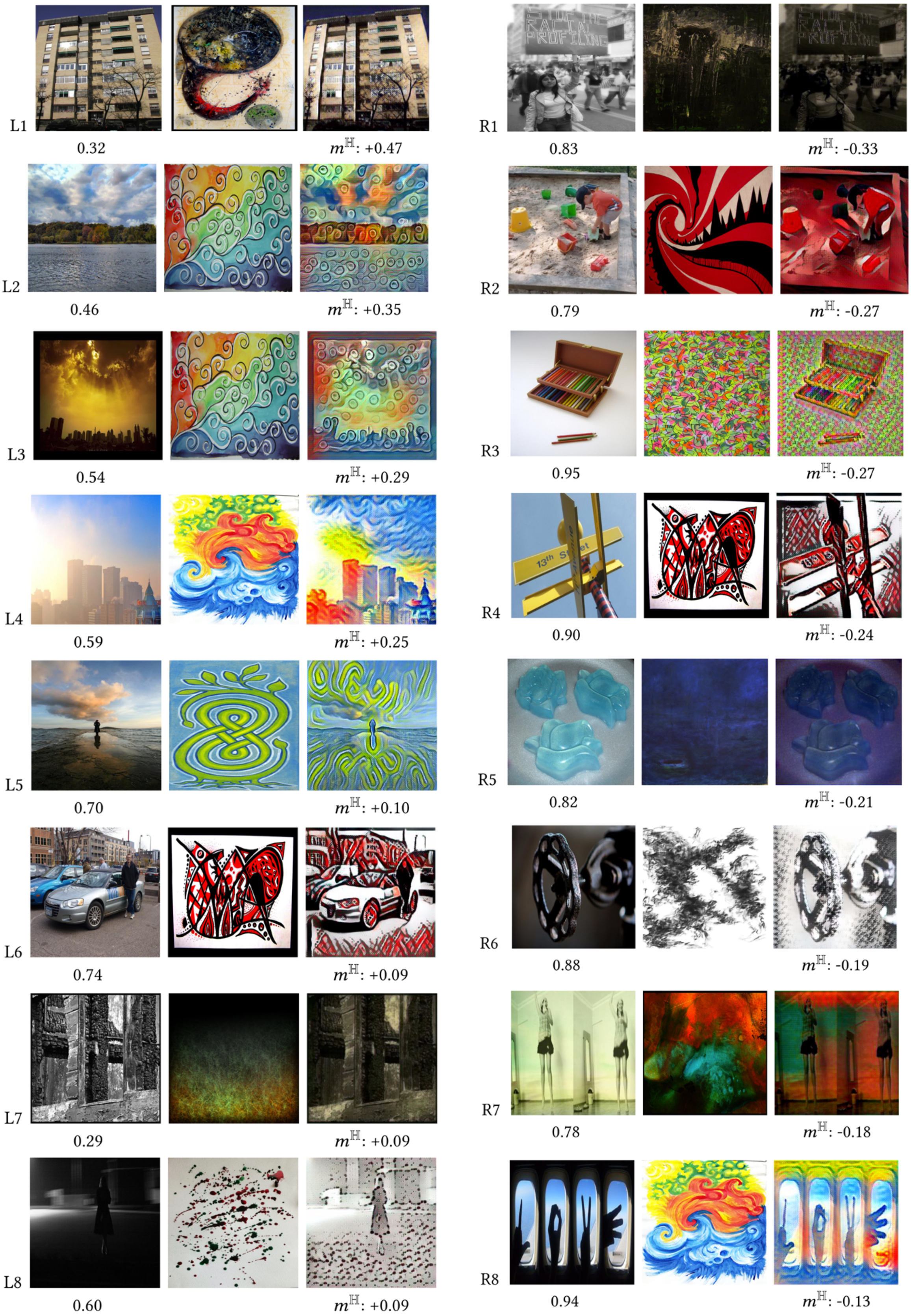


Fig. 4.6: *Sample results of the user study*

4.6.2. Increasing Image Memorability

In the third study we conduct a qualitative analysis on the images considered in our user study. Figure 4.6 reports some sample image-seed pairs for which we measured the largest memorability increases (L1-L8) and decreases (R1-R8).

The collected memorability scores and memorability gaps $m^{\mathbb{H}}$ are reported below the original and the stylized image. In the cases of Figure 4.6. L2-L5, typical non distinctive urban and outdoor scenes are transformed into a corresponding more colorful and joyful version. These new stylized versions of the images are probably perceived with a higher valence, arousal and, definitely, lower naturalness.

In the case of Figure 4.6. L6, the stylization clearly introduces complexity to the image. Still, the main objects in the scene (i.e., a man next to a car) are clearly visible.

Similarly, in the case of Figure 4.6. L8, the stylization process highlights the presence of a person in the scene, thus probably increasing its memorability. The samples in Figure 4.6. (right) illustrates that only increasing strangeness is not a sufficient condition for increasing image memorability. The drop in memorability in most of these cases may be explained with the fact that stylization diminishes the "readability" of the image, thus reducing the possibility for the observer to decode and retain the visual information observed in only one (1) second. For example, the text on the sign in the stylized versions of Figure 4.6. R1 is no longer readable, while in Figure 4.6. R2 the child in the playground is almost no longer recognizable. In these cases, the stylization process reduces the semantic information of the image, thus diminishing its memorability. Similarly, in Figure 4.6. R8, the stylization introduces colorful elements but makes the objects in the scene harder to recognize. In Figure 4.6. R6, the stylized image is a sort of sketch of the original one. In this case a possible explanation for the memorability decrease is the fact that the black and white style makes images not particularly distinctive and hard to remember.

Finally, it is interesting to observe that the same style applied to different images can induce opposite effects in terms of memorability variations. This confirms the validity of our framework where we select the best style for each given image, as there can be no universal style that is effective in memorializing all images. The samples in Figure 4.7. correspond to the images obtained with ***S-cube***. Specifically, we reported sample results where the optimal degree of stylization is automatically found to be $\alpha = 0.5$, 2 or 10, respectively for the top, central and bottom rows of Figure 4.7. The memorability increase is achieved by modifying the style of the images while retaining their original high-level content.

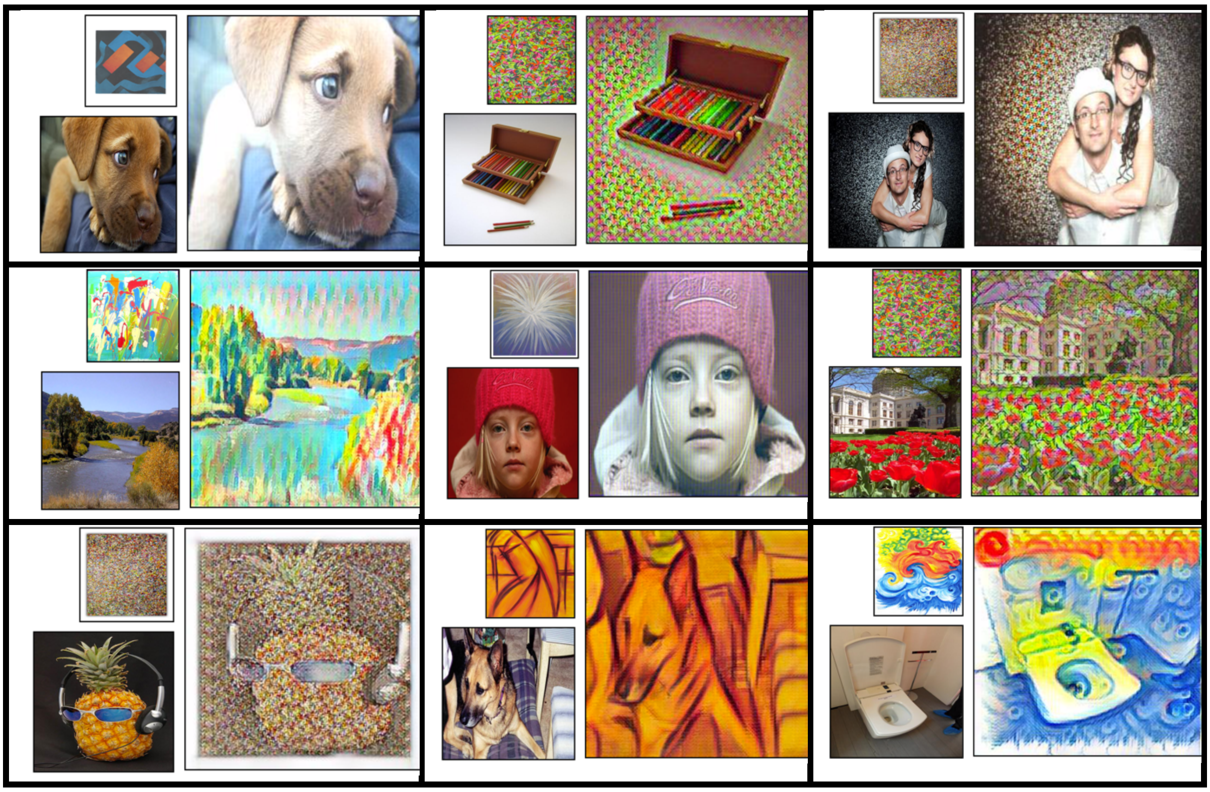


Fig 4.7: *Sample images obtained using* ***S-cube****: (top) α = 0.5, (central row) α = 2 and (bottom) α = 10.*

Chapter 5

CONCLUSIONS

Memories are in the center of our individuality. And although every human being has its own unique set of individual experiences that make up their memories, people tend to remember and forget the same images. This happens because images themselves differ in their memorability – a predictive value of whether an image is likely to be later remembered or forgotten.

During this process of learning and memorizing something, our brain forms connections called synapsis, between the neurons and the brain. Memories occur when we reactivate specific groups of neurons. It should be, however, noted that cases in which people show the enhancement of memory function in terms of recall known as hypermnesia, or the amnestic syndrome people who have difficulty forming new memories are excluded from this observations and conclusion.

A complex phenomenon of visual memorability depends on multiple factors. We presented recent large-scale visual memory studies that have shown that people have a remarkable ability at remembering specific details of images, with some images still being consistently more memorable or forgettable than others. The intrinsic property of memorability is, however, independent of the observers' individual differences, past experiences and it is reproducible across a population.

In general, images with people in them are the most memorable, followed by images of close-ups of objects. Least memorable are natural landscapes, although those can be memorable if they feature an unexpected element. Understanding this makes our task even more challenging: increasing an image's memorability without varying its high-level content.

By varying and changing the style of an image, we revealed and demonstrated how the probability of remembering an image could be increased. More precisely, we created and presented a novel approach to increase image memorability based on editing-by-filtering paradigm. In detail, we proposed a deep learning framework made by three components, namely the Scorer, the Synthesizer and the Selector.

The unique contribution of this approach is in the fact that, given an input image, the Selector automatically computes the style which guarantees the highest increase of that particular image's memorability which is, then, provided to the Synthesizer for generating a stylized version of the original image. The effectiveness of our approach, both in increasing memorability and in selecting the top memoralizable styles has also been evaluated on a public benchmark.

This is something no research has shown before and it is useful to other researchers in this field and beyond, which is why our efforts have been recognized by the ACM TOMM Editorial Board (Transactions on Multimedia Computing, Communications, and Applications, formerly TOMCCAP) as the most significant work in 2020 and awarded with Nicolas D. Georganas Best Paper Award [91].

Although our focus was on memorability, the suggested approach is extremely flexible, thus, allowing the possible applications to expand. For instance, by replacing the deep network implementing the Scorer, it can potentially be applied to other perceptual attributes, such as aesthetic quality or evoked emotions and similar.

Upcoming works will be dedicated to the extension of the proposed framework, following this direction. However, some other possible extensions for this work are the formulation of an approach for computing the degree of stylization α within a continuous range and designing a deep architecture where the style seeds are directly provided as input data to the Selector, instead of considering only the memorability gaps. Our expectations are that these modifications would lead to improved performance.

There are many exciting directions in which these findings could continue, from designing social media tools able to reduce human attempts and time of intervention in image modification activities, to coming up with technological solutions to automatically modify images and videos according to specific properties or attributes. This research has not only a commercial impact in marketing and fashion, but also in the way visual content influences people. Moreover, it has a profound impact in education and more generally in knowledge dissemination.

While the literature on predicting subjective attributes like memorability from visual data is extensive, few works have focused on the problem of synthesizing images to modify these properties. In this work we considered the neural style transfer procedure and techniques. Nevertheless, we believe that other deep learning models, such as Generative Adversarial Networks (GANs), could be exploited for this purpose and our future works will explore this direction. Further research may also include additional psychophysical experiments to both improve the reliability of results and collect more experimental data.

Due to the fact that the key drivers of the variations in memorability of images are not always and completely reasonable and logical, large interest in cognitive science to understand some of these complex principles has been noted. By investing some of images' cognitive properties, among which also memorability, a framework that uses GANs could contribute to defining what each of the attributes involves and stands for, along with providing more natural-looking images. In future work we will also expand the behavioral experiment component with the purpose of discovering more precisely the way in which image manipulations and image properties variations are affecting human memory performance.

Appendix A

PRODUŽENI APSTRAKT NA SRPSKOM JEZIKU

Fotografije su jedan od najdominantnijih tipova medija koji se svakodnevno u milijardama pojedinačnih primera prenose na online platforme različitih tipova. Srazmerno tome se i broj platformi za deljenje ovog sadržaja neprestano uvećava, a posledično raste i interesovanje za razumevanjem sadržaja kompletne fotografije ili nekih njenih delova, kao i stila, ali i emocija koje one izazivaju. Ovo istraživanje fokusirano je ka pronalasku rešenja za problem automatskog povećanja pamtljivosti jedne arbitrarno izabrane fotografije.

Znajući da svaka osoba ima jedinstven skup individualnih iskustava na osnovu kojih evocira događaje i sećanja, očekivalo bi se da ove individualne razlike utiču na odluku prilikom memorisanja određenog sadržaja. Iznenađujuće, istraživanja su pokazala da ljudi imaju tendenciju da pamte i zaboravljaju iste stvari, što je uzrokovano suštinskim karakteristikama same slike koje mogu ukazati na to koliko će određena slika biti lako upamćena od strane različitog broja posmatrača. Pamtljivost slike je, prema tome, objektivna i kvantitativna mera same slike, nezavisna od posmatrača, što je ujedno i preduslov za dalje manipulacije i računarska predviđanja.

Radovi iz oblasti Računarskog vida i Multimedija pokazali su da se uz pomoć korišćenja modela dubokog učenja (eng. *Deep Learning*) pojedina svojstva slike mogu automatski predvideti. Ovo je podstaklo definisanje još izazovnijeg istraživačkog pitanja „Može li se transformisati proizvoljna fotografija na način da ona bude lakše upamćena?". Za adekvatnu formulaciju ovog pitanja, potrebno je poći od definisanja mera pamtljivosti, ali svakako bi ovakve metode za automatsko povećanje ovog svojstva fotografije mogle imati uticaja u mnogim oblastima, od obrazovanja, do primene u gejming industriji ili prilikom izrade reklamnog sadržaja.

Za bavljenje ovom problematikom, izabrani pristup inspirisan je paradigmom uređivanja fotografija primenom filtera, usvojenom u aplikacijama Instagrama (eng. *Instagram*) i Prizme (eng. *Prisma*), sa razlikom u tome što korisnici dveju aplikacija moraju da prolaze kroz listu dostupnih filtera pre nego što pronađu željeno rešenje za modifikaciju ulazne fotografije, što sam postupak uređivanja čini resursno i vremenski neefikasnim. Na kraju, ishod ovakve modifikacije ulazne fotografije ne garantuje da će aplikacijom određenog filtera i promenom stila originalne slike izlazna fotografija biti pamtljivija. U ovom radu preokrećemo proces tako da u zavisnosti od karakteristika ulazne fotografije (korisniku) stiže predlog u vidu skupa stilova čijom bi se aplikacijom na ulaznu sliku putem neuronskog algoritma prenosa stila pružila mogućnost povećanja pamtljivosti date fotografije. Kao rezultat ovog procesa stiže dokaz o postojanju mogućnosti za automatizacijom čitavog procesa i pronalaskom stila koji najviše odgovara datoj fotografiji, čime se, posledično smanjuje i broj ljudskih pokušaja potrebnih za pronalazak najboljeg podudaranja.

Dalje, demonstrirana je efikasnost predloženog pristupa eksperimentalnom validacijom na javno dostupnom skupu podataka LaMem, uz izvođenje kvantitativne procene, kao analize iskustva korisnika. Radi demonstracije fleksibilnosti predloženog okvira, analiziran je i uticaj različitih izbora implementacije, uz prikaz nekoliko kvalitativnih rezultata koji pružaju dodatne uvide o vezi između stila fotografije i njene pamtljivosti.

Predloženi pristup se oslanja na napredak u oblasti sinteze slika i usvaja duboku arhitekturu za generisanje pamtljive slike, na bazi date ulaz ne slike i stilskih setova slika. Automatski izbor najboljeg stila, odnosno stila koji će u najvećoj meri povećati pamtljivost ulazne slike, oslanja se na duboke modele i predlaže se novo rešenje zasnovano na ovom učenju.

A.1. Predmet i ciljevi istraživanja

Korisnici danas generišu i dele stalno rastuću količinu multimedijalnog sadržaja, poput fotografija i video zapisa. Ilustracije radi, u 2017 godini je video činio 74% sadržaja koji je prenesen Internetom, a na popularnoj platformi za deljenje slika Instagram se dnevno generiše preko 95 miliona fotografija i video zapisa. Osim niza pozitivnih efekata informaciono-komunikacionih tehnologija, brz protok podataka neminovno je doveo i do preopterećenosti sadržajem kojim su ljudi svakodnevno izloženi. U ovakvom okruženju, pitanje mogućnosti zadržavanja informacija prezentovanih na ovaj način i njihove ponovne reprodukcije, odnosno pamtljivosti svih tih sadržaja, dobija na naučnom značaju.

Polazeći od idioma prema kom ”slika vredi hiljadu reči”, aluzije na to da se složena ideja može objasniti samo jednom slikom, inspirisan je nastanak informacijske grafike tzv. infografike. Ovim oblikom vizuelne prezentacije informacija, podataka i znanja, grafički dizajneri i marketinški stručnjaci kreirali su mogućnost prenosa kompleksnih informacija brže i jasnije. Ipak, život u instant društvu u kom je imperativ svake aktivnosti brzina i u kom količina vizualnih stimulusa kojima su ljudi izloženi nastavlja da raste, otvara niz pitanja o kapacitetu skladištenja svih tih informacija u ljudskom mozgu, kao i o kriterijumu selekcije koji ljuski mozak koristi kako bi zadržao sve te informacije. Ono što je primećeno u dosadašnjim istraživanjima je tendencija mozga da apsorbuje i zadržava znanja predstavljena vizuelno, kroz video zapise i fotografije, favorizujući ovaj tip signala u odnosu na sve druge. Međutim, ovaj efekat superiornosti nije primenljiv na situaciju u kojoj se želi uporediti sadržaj dve fotografije i slučajeve u kojima se neke slike lako pamte, dok se druge brzo zaboravljaju. Na kraju, pored značajnog napretka u ovoj oblasti, automatska transformacija slike sa ciljem povećanja šansi da ona bude upamćena predstavlja još uvek otvoreno istraživačko pitanje. Pronalazak modaliteta za privlačenje i zadržavanje pažnje subjekta, uz istovremeni rast verovatnoće za ponovno prepoznavanje jednom viđenog sadržaja (pamtljivosti) mogao bi imati niz praktičnih primena u mnogim disciplinama.

Radi boljeg razumevanja aktuelnog stanja u oblasti, prilikom koncipiranja ovog istraživanja analizirana su najaktuelnija stanja u nekoliko oblasti, poput kognitivne neuronauke, psihologije, gde je primećen porast interesovanja za ispitivanjem raznih medijskih svojstava, konkretno boje fotografije, kvaliteta, transparentnosti, ali i svojstva koja utiču na način percepcije same fotografije, poput stepena njene pamtljivosti. Za oblast marketinga je ovo od posebne važnosti budući da se faktor iznenađenja često koristi sa ciljem pobuđivanja pažnje i podsticanja potrošačkog impulsa krajnjeg korisnika, a idealno i aktivira njihova motivacija i spremnost da kupe određeni proizvod sa date fotografije. Aplikacije koje su analizirane u ovom radu nastale su upravo kao posledica sadejstva marketinga i grafičkog dizajna, a daljom identifikacijom obrazaca ponašanja ljudi i njihovih navika i percepcije, izdvajaju se znanja korisna za niz drugih primena, recimo u studijskom materijalu za učenje i trening, ukoliko je poznato da kognitivna prezasićenost nastaje upravo usled kompleksnosti materijala, faktora koji remeti proces pamćenja i učenja. Na sličan način bi se mogli izabrati ne samo delovi slika i studijskog materijala, već i segmenti jednog video zapisa koji imaju najveći potencijal pamtljivosti.

U okviru ove disertacije istraženi su načini na koje se automatski može povećati efikasnost apsorpcije informacija prezentovanih u vizuelnom obliku (fotografija) manipulacijom stila jedne fotografije. Takođe, istražene su mogućnosti korišćenja sve veće količine javno dostupnih baza fotografija, sa idejom da se pronađu novi modeli mašinskog učenja koji bi se primenili za povećanja pamtljivosti sadržaja. Ispitana je i mogućnost objektivnog i automatskog merenja pamtljivosti koja je validirana u stvarnoj interakciji korisnika sa računarom.

Prethodnih godina u nizu publikovanih naučnih radova pružen niz dokaza o tome da je pamtljivost intrinzično svojstvo slike. Ovo znači da različiti posmatrači iste fotografije pokazuju približno iste performanse u procesu pamćenja i zaboravljanja, što je dovelo u pitanje ranije dominantne stavove u nauci gde je istican značaj individualnih razlika. Napredak u istraživanjima iz domena računarskog vida doveo je ne samo do mogućnosti merenja pamtljivosti slike, već i do rešenja koja su demonstrirala njeno automatsko predviđanje posredstvom naprednih modela dubokog učenja obezbeđujući tačnost približnu performansama ljudi. Ovi pionirski koraci u istraživanju i razvoju metoda za analizu pamtljivosti slike otvorili su mogućnost daljeg kombinovanja podataka, tj. stila fotografije poput boja i teksture, sa njenim osnovnim sadržajem u cilju izrade modela prema kom će se pamtljivost moći izračunati, predvideti a zatim i automatski povećati.

Osnovni cilj je svakako razvoj modela za automatsko povećanje pamtljivosti proizvoljne fotografije, uz nekoliko ključnih doprinosa:

- istraživanje mogućnosti u vezi sa povećanjem pamtljivosti slike, uz zadržavanje sadržaja visokog nivoa, drugim rečima, modifikujući samo stil fotografije;

- analiza problema sinteze slike zasnovanog na stilu, u kontekstu dubokih arhitektura i predlog automatskog metoda za povratak predloženog skupa stilova za koje se očekuje da bi dovele do najvećeg povećanja pamtljivosti ulazne slike;

- predlog rešenja za trening mreže (tzv. *Selector*) koja dozvoljava efikasno učenje modela sa redukovanim brojem podataka za trening, a uz relativno velike varijacije stila slike.

Prikaz aktuelnog stanja u oblasti dat je sa aspekta različitih naučnih disciplina, uz zadržavanje fokusa na tri glavne linije istraživanja usmerene ekskluzivno ka pamtljivosti, redom, studije koje adresiraju automatsku manipulaciju slika uz izmenu njenih perceptivnih atributa, zatim studije koje analiziraju vizuelnu pamtljivost i one koje su istakle nalaze radova baziranih na transferu neuralnog stila.

A.2. Korišćeni alati

Prilikom koncipiranja ovog istraživanja napravljen je iskorak iz dosadašnje analize i merenja pamtljivosti slika i za cilj istraživanja postavljeno kreiranje modela za automatsko povećanje pamtljivosti proizvoljno izabrane slike. Prema tome, ideja je i pružanje doprinosa boljem razumevanju svih aspekata u vezi sa temom, od same ljudske memorije i njenog potencijala za skladištenjem informacija, ali i za evociranjem već viđenog sadržaja, preko razumevanja svojstava vizuelnih sadržaja, segmentacije slike i kombinovanja njene teksture, boja i ostalih elemenata koji utiču na percepciju same fotografije. Do ovoga se dolazi najpre razumevanjem postojećih mera pamtljivosti sa ciljem objektivne procene pamtljivosti celokupne fotografije, a zatim i analizom već razvijenih alata koji svakako doprinose preciznijem definisanju odnosa same pamtljivosti i drugih svojstava slike.

Pristup rešavanju ovog problema svakako se u najvećoj meri oslanja na studije iz domena računarskog vida gde se do interpretacije i boljeg razumevanja jedne fotografije dolazi nizom tehnika koje tipično uključuju prepoznavanje, identifikaciju i lokalizaciju objekata korišćenjem neuronskih mreža, klasifikaciju i segmentaciju slike ili njenih delova i slično. Ipak, stilizacija slike vezana za transfer neuralnog stila (eng. *Neural Style Transfer*, NST), u najdirektnijoj je vezi sa istraživanjem.

Zbog svega ovoga, pristup je zasnovan na primeni tehnika računarskog vida i veštačke inteligencije na čijem se preseku nalazi potencijal za razvoj inovativnog modela sa mnoštvom daljih primena među kojima su:

- identifikacija sadržaja koji ima najveći potencijal da bude upamćen;
- otkrivanje navika i načina na koji ljudi percipiraju multimedijalni sadržaj;
- podrška boljem procesu donošenju poslovnih odluka primenom veštačke inteligencije.

Za realizaciju ovog cilja pristupilo se korišćenju relevantnih naučnih metoda za prikupljanje, obradu, prikaz i analizu podataka, a za analizu eksperimentalnih rezultata primenjene su odgovarajuće statističke tehnike. Za objektivnu procenu i samo povećanje stepena pamtljivosti slika korišćena su programska okruženja otvorenog koda Tenzorflou (engl. *TensorFlow*) i Pajtorč (engl. *PyTorch*), kao i programski jezik Pajton (engl. *Python*).

A.3. Korišćeni skupovi podataka

U ovom poglavlju biće prikazan način izbora, veličina, kao i konstrukcija uzorka i korišćenih setova podataka u ovom istraživanju.

Mali skupovi podataka prethodno korišćeni za neke osnovne zadatke prepoznavanja nisu bili najpogodniji za istraživački zahtev definisan već početnom inspiracijom idejom za razvoj modela pogodnog za arbitrarne slike koje imaju mnogo više varijabilnosti od onih dostupnih u manjim skupovima podataka. Dostupnost velikih skupova podataka omogućila je upotrebu moćnih modela dubokog učenja prilikom zadataka prepoznavanja objekata, a progresivan razvoj je uočen u istraživačkoj grupi Univerziteta u Torontu zbog čega smo odlučili da primenu modela nađemo najpre u AlexNet arhitekturi, pre-treniranoj za dva tipa zadataka i na dva različita tipa podataka, a sačinjene od ogromnog broja označenih slika.

AlexNet arhitekturu čini osam slojeva, od kojih pet konvolutivnih i tri potpuno međusobno povezana sloja. Pošavši od činjenice da brojni istraživači koriste hibrid već predloženih arhitektura sa ciljem poboljšanja performansi konvolucionih neuronskih mreža (eng.

Convolutional Neural Network, CNN), u slučaju ovog istraživanja, izbor hibridnih CNN smatra se prikladnijim upravo zbog arbitrarnog izbora slika, imajući u vidu činjenicu da je mreža prethodno bila trenirana za prepoznavanje objekata. Druga korišćena arhitektura je VGG16, za zadatke klasifikacije i detekcije, prethodno predložena od strane istraživača sa Oksforda, uz napomenu da model postiže preciznost od 92.7%, čime je predstavljena prednost u odnosu na AlexNet. ImageNet projekat je započet 2009. godine i do sada je inspirisalo rast interesovanja u ovom polju istraživanja, upravo prilikom izraženih potreba za većim brojem podataka. Stoga je u ovom slučaju korišćena za pred-trening Hybrid-CNN za zadatke klasifikacije objekata, dok je za scene korišćen set podataka zvan *Places* (mesta, u doslovnom prevodu), zasnovan na principima ljudske percepcije i sačinjenom od 10 miliona fotografija različitih mesta i scena.

Efikasnost modela razvijenog za automatsko unapređenje pamtljivosti slika evaluirana je eksperimentalno na javno dostupnoj bazi od oko 60.000 fotografija, preciznije 58.741 slika iz LaMem baze podataka. Dodatno je za potrebe ovog istraživanja i zbog ideje da se izbegne modifikacija sadržaja visokog nivoa, u obzir uzeta i baza DeviantArt sačinjena od 500 apstraktnih slika. Motivacija je krajnje praktična, budući da apstraktne slike ne sadrže mnogo informacija, osim kombinacije teksture i boja, što ih čini idealnim upravo za automatski proces modifikacije karakteristika i elemenata niskog nivoa.

LaMem baza podataka predstavlja najveći i najaktuelniji postojeći skup anotiranih fotografija u ovom istraživačkom domenu, konstruisan od strane istraživača sa Masačusets Instituta za tehnologiju sa ciljem da doprinesu istraživačkoj zajednici u napretku prilikom procene vizuelne pamtljivosti. Skup je sačinjen od nekoliko prethodno postojećih skupova fotografija i apstraktnih slika.

DeviantArt (dA) set apstraktnih slika nastao je prikupljanjem sa online mreže i najveće artističke zajednice sa oko 360 miliona umetničkih dela i sa oko 35 miliona registrovanih korisnika, među kojima su i amateri i profesionalni umetnici. Sva ova umetnička dela organizovana su i podeljena u kategorije npr. fotografije, digitalna umetnost, tradicionalna umetnost itd. Kategorija izabrana za potrebe ovog istraživanja je, kako je već spomenuto, apstraktna umetnost.

Poslednja dva skupa podataka ocenjena su kao najprikladnija za potrebe istraživanja, a korišćen je set od 45.000 slika podeljenih, zatim, nasumično u dva odvojena poskupa od po 22.500 slika. Analizom odgovora CNN slojeva visokog nivoa moglo se shvatiti koji su predmeti i regioni slike bili u pozitivnoj i negativnoj korelaciji sa pamtljivošću, što je upravo i nalaz koji je pružio konkretan metoda za dalju manipulaciju pamtljivosti slike.

Pored kvantitativne evaluacije, primenom igre memorije vršena je procena kvalitativnog aspekta rezultata predloženog modela, sa ciljem prevazilaženja limitacija prethodno objavljivanih radova na ovu temu koji su se oslanjali na analize malih baza podataka i uglavnom veoma specifičnog tipa fotografija. Varijacije do kojih dolazi prilikom analize velikog skupa podataka, posebno su prikladne za analizu performansi predloženog modela. Učesnici kvalitativne studije bili su mahom student, a cilj je bio da se prikupi oko 150-200 odgovora po slici, odnosno toliki broj polno ujednačenih ispitanika, što je i ostvareno.

A.4. Metodologija

U okviru rada za potrebe ove teze razvijen je alat za automatsko povećanje pamtljivosti fotografija promenom i modifikacijom njihovog stila, to jest manipulacijom karakteristika nižeg nivoa (engl. *low-level features*), uz očuvanje reprezentacija višeg nivoa (engl. *high-level features*).

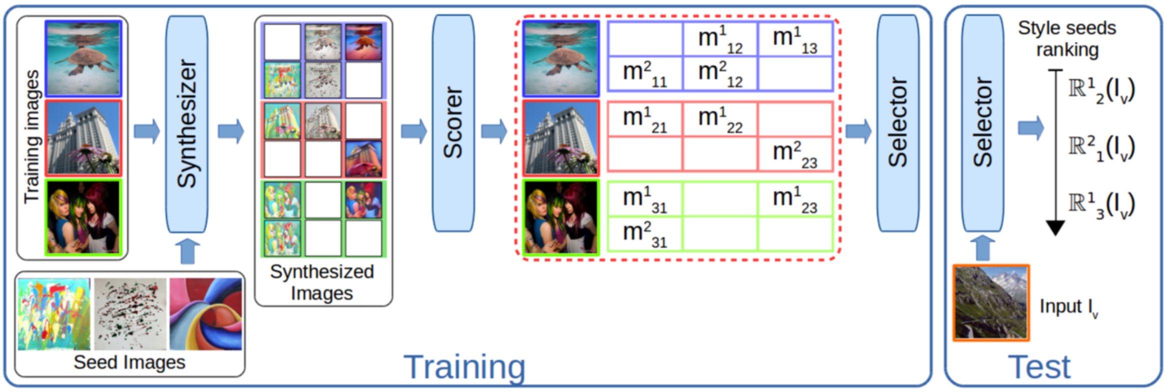
Metodološki okvir za automatsko povećanje pamtljivosti arbitrarne ulazne fotografije koncipiran je na postulatima efikasnosti uz upotrebu metoda prenosa neuronskog stila kako bi se stvorile stilizovane slike koje ipak uspevaju da zadrže sadržaj visokog nivoa same slike. Polazna osnova i jedna od inspiracija za razvoj nove metodologije, kao i inovativnog alata koji omogućava automatizaciju kompletnog procesa dolaze od grupe istraživača sa Masačusets Instituta za tehnologiju (MIT) iz Laboratorije za veštačku inteligenciju koja je razvila metodu za modifikovanje pamtljivosti fotografija ljudskog lica uz zadržavanje identiteta osoba i drugih facijalnih karakteristika poput godina starosti, atraktivnosti i mapiranog emocionalnog intenziteta (tzv. emocionalne magnitude).

Predloženi pristup koristi se metodom prenosa neuralnog stila za stvaranje stilizovanih slika, uz neophodnost implementacije brojnih modifikacija kako bi se osiguralo povećanje pamtljivosti slika, uz istovremeno očuvanje originalnog stila i sadržaja slike.

Okvirno, metod je dizajniran uz stavljanje naglaska na performanse povećanja pamtljivosti slike. Pristup predložen u metodološkom delu ove disertacije artikulisan je kroz tri glavne komponente:

1) *Selektor stila (The Style Selector)*
2) *Skorer (The Scorer)*
3) *Sintetizator (The Synthesizer)*

Zbog ovoga je pristup nazvan ***S-Cube*** ili **S^3**, a radi jednostavnijeg razumevanja ideje, sledi ilustracija koncepta i pregled modela razvijenog za potrebe ovog istraživanja.



Ilustracija modela razvijenog za potrebe istraživanja

Sa fotografije se vide dve osnovne faze – trening i test faza. U vreme treninga, Sintetizator (S) i Skorer (na ilustraciji označen sa “M”) služe za generisanje podataka o treningu za Selektor stila slike (označen kao “R” na ilustraciji). Podaci o treningu su istaknuti u delu uokvirenom crvenom isprekidanom linijom. U vreme testiranja, Selektor stila slike za svaku novu sliku nudi sortiranu listu stilova, zasnovanu na predviđenom povećanju skora memorabilnosti slike.

Selektor je prema tome jezgro predloženog pristupa i njegov zadatak je da za arbitrarnu ulaznu sliku i set skupova stilskih slika preuzme podskup slika koji će moći da proizvede najveći porast pamtljivosti ulazne slike. Drugim rečima, Selektor stila slike predviđa očekivano povećanje ili smanjenje pamtljivosti koje će svaki stil proizvesti na ulaznoj arbitrarnoj slici, a shodno tome i rangira skup prema očekivanom povećanju pamtljivosti. Tokom treninga, Sintetizator i Skorer se koriste za generisanje slika iz mnogih parova ulaznih slika. Svaka ulazna slika je zatim povezana sa relativnim povećanjem ili smanjenjem pamtljivosti koja se dobija sa svakim od stilova. Pomoću ovih informacija možemo naučiti da predvidimo povećanje ili smanjenje pamtljivosti za nove ulazne slike, te stoga rangiranje vršimo prema očekivanom povećanju.

Istraživanje demonstrira i mogućnost predviđanja ne samo stilova koji imaju najveći potencijal pamtljivosti, već i za svaki stil automatsko izračunavanje optimalnog stepena stilizacije.

Uz sve ovo, primenjena je i kvalitativna analiza za ispitivanje korelacije između pamtljivosti i ostalih opažajnih kvaliteta poput afektivnog kvaliteta, zanimljivosti slike i sl.

Uzimajući u obzir multidisciplinarni karakter ovog istraživanja, u metodološkom domenu ono svakako predstavlja kombinaciju kvalitativne i kvantitativne analize. Zbog naglaska na dubljem razumevanju i elaboraciji slika, kvalitativna metoda smatra se korisnom za područje socijalnog istraživanja, odnosno tog aspekta u celokupnoj izradi disertacije. Ovo empirijsko kvalitativno proučavanje doprinosi sakupljanju važnih činjenica koje bi mogle biti propuštene opštim kvantitativnim proučavanjem. Sa druge strane, kvantitativni pristup daju bolji uvid u ustanovljene uzročno-posledične odnose između ispitivanih komponenata.

Na kraju, realizovana je i eksperimentalna evaluacija efektivnosti razvijenih modela posredstvom korisničke studije kroz strukturiranu igru memorije. Evaluacija je realizovana elektronski podržanim putem, slanjem direktnog linka do igrice, posredstvom Interneta.

Trening faza

Tokom trening faze vrši se učenje tri modela. Sintetizer (Synthesizer, Sy) i skorer (Scorer, Sc) se koriste za generisanje slika iz velikog broja input parova: slika-semenka stila. Posledično, uče i da skoruju ove parove. Drugim rečima, svaka input slika je zatim povezana sa relativnim povećanjem, odnosno smanjenjem skora pamtljivosti dobijenim u kombinaciji sa svakim „semenom" stila. Zahvaljujući ovim informacijama možemo da predviđamo porast odnosno smanjenje skora pamtljivosti nove slike, a zatim i da rangiramo sve stilove na osnovu očekivanog porasta pamtljivosti, što je i cilj. U nastavku sledi detaljniji opis ovih aktivnosti.

Model skorovanja, tzv. Scoring model (Sc), vraća vrednost pamtljivosti $Sc(\mathbf{I})$ generičke slike $\mathbf{I}$, što se uči pomoću trening seta slika anotiranih pripadajućim skorom pamtljivosti:

$$\mathcal{M} = \{I_i^M, m_i\}_{i=1}^{I}$$

Osim trening seta, u obzir se uzima i razmatra generišući set prirodnih slika:

$$\mathcal{G} = \{I_g^G\}_{g=1}^{G}$$

kao i set stilskih slika:

$$\mathcal{S} = \{S_s\}_{s=1}^{S}$$

Zatim se procesom sintetizacije posredstvom tzv. Synthesizer-a proizvodi slika na osnovu para slike-semenke stila,

$$I_{gs} = Sy(\boldsymbol{I_g^G}, \boldsymbol{S_s})$$

Scoring model *Sc* i Synthesizer *Sy* su neophodni koraci za trening selektora "semenki" stila, Selector *Sl*. Za svaku $I_g^G \in \mathcal{G}$ sliku i za svaki stil $S_s \in \mathcal{S}$ postupkom sinteze stvara se I_{gs}. Scoring model se koristi za računanje međuprostornog jaza (praznina) između skorova pamtljivosti sintetizovane i originalne slike:

$$m_{gs}^{Sc} = Sc(\boldsymbol{I_{gs}}) - Sc(\boldsymbol{I_g^G})$$

Nadalje se koncentracija svih skorova vezanih za „semenke“ stila koristi za učenje selektora tih semenki, a za potrebe ovoga je konstruisan trening set prirodnih slika koje su označene sa svim spomenutim razlikama u skorovima pamtljivosti:

$$\mathcal{R} = \{I_g^G, m_g^{Sc}\}_{g=1}^{G}$$

Proces selekcije “semenki” zapravo je definisan kao problem regresije i mapiranja **R** između same slike i svih pridruženih vektora rezultata međuprostornog jaza skorova pamtljivosti. Od momenta od kog je selektor treniran na **R**, u stanju je da proceni vektor svih međuprostornih praznina za test sliku, što je daleko brže od pokretanja Synthesizer Scorera. Uz ovo je obezbeđeno i rangiranje “semenki” stilova u zavisnosti od njihovih mogućnosti da fotografiju učine pamtljivijom, odnosno, dobija se pregled najboljih semenki koje odgovaraju najvećem porastu skora pamtljivosti.

Test faza

Tokom test faze i zadate nove slike $\mathbf{I}_v$, selektor “semenki” stilova je angažovan za predviđanje vektora međuprostornih skorova pamtljivosti za sve stilove: $\mathbf{m}_v = Sl\ (\mathbf{I}_v)$. Rangiranje stilova je zatim izvedeno iz vektora $\mathbf{m}_v$, a na bazi ovih rangova, Synthesizer je apliciran na test sliku $\mathbf{I}_v$ uzimajući u obzir isključivo Q semeke stila $\mathbf{S}_s$ proizvodeći set stilizovanih slika:

$$\{I_{qs}\}_{q=1}^{Q}.$$

A.4.1. Eksperimentalna validacija

Predloženi metod nazvan ***S-cube*** eksperimentalno je validiran korišćenjem prethodno opisanog LaMem skupa podataka. Tokom ove faze, u obzir su uzete:

- **45,000 slika za trening**
- **10,000 slika za test**
- **3,741 slika za validaciju.**

Prvim postupkom izvršena je podela LaMem trening seta u dva podskupa, svaki od po 22,500 slika, $\mathcal{M}$ i $\boldsymbol{E}$, korišćenih za treniranje prediktora, čije su se performance evaluirale računanjem rang korelacije između predviđene i stvarne pamtljivosti na LaMem test setu. Dobijeni skor rang korelacije iznosi 0.63 za oba modela, dok je skor 0.64 dobijen na ukupnom LaMem trening setu, što je slično ljudskom učinku od 0.68. Test set LaMem slika sačinjen od 10,000 fotografija korišćen je za učenje selektora stilova. Podelom LaMem test seta na test za trening, validaciju i testiranje za selector, vršeno je u odnosu 8:1:1, drugim rečima 8,000 slika za trening, a po 1,000 za validaciju i testiranje.

Kada su u pitanju "semenke" stilova, pamtljivost smo procenjivali i na osnovu svih slika iz DeviantArt baze, izdvajanjem 50 najpamtljivijih i 50 najmanje pamtljivih slika iz ovog seta podataka. Ovo je dalo rezultat pamtljivosti u rangu između 0.556 i 0.938.

Sve ovo rađeno je sa ciljem da se samom istraživačkom postupku pridruži dokaz koji potvrđuje:

- da predložen metodološki okvir poseduje kapacitet povećanja pamtljivosti proizovljnih slika,
- ali i da demonstrira da je eksterni prediktor (*Ev*) dobar posrednik za procenu korisnika.

Kompletna studija predstavlja prvu dubinsku analizu koja uspeva da postigne automatsko povećanje pamtljivosti arbitrarno izabrane slike, što je upravo razlog zbog kog direktno poređenje sa prethodnim studijama, niti ovakve vrste kvantitativnih analogija nisu izvodljive. Aktuelno stanje u oblasti uspelo je da otvori put ovakvim i sličnim aplikacijama i raznim drugim vrstama manipulacije slikom, uz evidentno moguće precizno računanje mapa pamtljivosti slika.

A.5. Rezultati

Interni i eksterni prediktori trenirani su na prirodnim slikama zbog čega je diskutabilno pitanje da li ishod ovih prediktora važi za stilizovane slike. Ipak, za procenu povezanosti spoljnog prediktora i rezultata ljudskog angažmana, računali smo vrednost Pirsonovog koeficijenta korelacije i MSE uzimajući u obzir automatski predviđene, ali i realne skorove pamtljivosti slike u setu za verifikaciju $\mathcal{Z}$. Vrednost Pirsonovog koeficijenta odgovara originalnoj i stilizovanoj slici u setu $\mathcal{Z}$ i iznosi $\varrho = 0.66$ za originalnu i $\varrho = 0.50$ za stilizovanu, uz napomenu da je ljudski performans $\varrho = 0.68$. Računanjem Spirmanovog koeficijenta dobijeni su isti zaključci kao i prilikom računanja Pirsonovog koeficijenta korelacije. Svi ovi rezultati svedoče o visokoj korelaciji između realnog i predviđenog skora pamtljivosti.

A.6. Zaključak

Složen fenomen vezan za vizuelnu memoriju zavisi od mnogobrojnih faktora. Variranjem i modifikovanjem stila slike, uspešno se otkriva način na koji se verovatnoća da će slika biti upamćena isto tako povećava. Shodno tome, kreiran je i predstavljen nov pristup povećanju pamtljivosti slike baziran na postulatima editovanja slike apliciranjem filtera.

Jedinstveni doprinos ovog pristupa sastoji se u procesu automatskog izračunavanja stila koji daje garanciju za povećanje pamtljivosti date i izabrane slike, generišući stilizovanu verziju originalne slike.

Na prethodnim stranicama prikazano je istraživanje kakvo do sada nije rađeno u ovoj istraživačkoj oblasti, a ni šire. Iako je fokus bio na pamtljivosti, predsloženi pristup je krajnje fleksibilan i ima mogućnost proširivanja polja aplikacija, recimo na druge perceptivne atribute poput estetskih, ali i afektivnog kvaliteta, evociranih emocija i slično.

Druga proširenja ovog rada mogla bi ići u smeru izmena prilikom računanja stepena stilizacije, ali i predviđanje subjektivnih atributa tehnikom prenosa neuronskog stila, uz uverenost da se i drugi modeli dubokog učenja mogu iskoristiti u ove svrhe, konkretno Generativne Adversarial Networks (GANs), te će predstojeći radovi svakako biti posvećeni proširivanju predložen og okvira prateći ovaj smer.

Bibliography

[1] Cisco, «Cisco Annual Internet Report (2018–2023) White Paper.,» 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html.

[2] A. F. Blackwell, «Correction: A Picture is Worth 84.1 Words.,» in *In C. Kann (Ed.), Proceedings of the First ESP Student Workshop , pp. 15-22.*, 1997.

[3] A. G. Goldstein and J. E. Chance, "Visual recognition memory for complex configurations.," *Perception & Psychophysics.,* vol. 9, p. 237–241, 1970.

[4] V. Mehrpour, Y. Mohsenzadeh, A. Jaegle, T. Meyer , O. Aude e N. C. Rust, «A neural correlate of image memorability in inferotemporal cortex.,» *Journal of Vision.,* vol. 19, n. 91c, 2019.

[5] M. I. Foster e M. T. Keane, «The Role of Surprise in Learning: Different Surprising Outcomes Affect Memorability Differentially.,» *Forthcoming Topic: The Ubiquity of Surprise: Developments in Theory, Converging Evidence, and Implications for Cognition.,* vol. 11, n. 1, pp. 75-78, 2018.

[6] T.-W. Chang, C. Kinshuk, Y. Nian-Shing e Pao-Ta, «The Effects of Presentation Method and Information Density on Visual Search Ability and Working Memory Load.,» *ACM Digital Library: Computers and Education.,* vol. 58, n. 2, pp. 207-222, 2012.

[7] S. Maren, «Long-term potentiation in the amygdala: a mechanism for emotional learning and memory.,» *Trends in Neurosciences.,* vol. 12, n. 2, pp. 561-567, 1999.

[8] L. G. Standing, «Learning 10,000 Pictures.,» *The Quarterly journal of experimental psychology.,* vol. 25, n. 2, 1973.

[9] A. K. Anderson, P. E. Wais e J. D. Gabrieli, «Emotion enhances remembrance of neutral events past.,» *Proceedings of the National Academy of Sciences of the United States of America.,* vol. 103, n. 5, pp. 1599-1604, 2006.

[10] P. Isola, D. Parikh, A. Torralba e O. Aude, «Understanding the Intrinsic Memorability of Images.,» *Advances in Neural Information Proceeding Systems 24 (NIPS),* 2011.

[11] Z. Bylinskii , P. Isola, C. Bainbridge , A. Torralba e O. Aude, «Intrinsic and extrinsic effects on image memorability.,» *Vision Research.,* vol. 116, pp. 165-178, 2015.

[12] A. Khosla, A. S. Raju, A. Torralba e O. Aude, «Understanding and Predicting Image Memorability at a Large Scale.,» in *International Conference on Computer Vision (IEEE)*, 2015.

[13] L. Cahill e J. L. McGaugh, «A novel demonstration of enhanced memory associated with emotional arousal.,» *Consciousness and Cognition: An International Journal.,* vol. 4, n. 4, p. 410–421, 1995.

[14] S. Russell e P. Norvig, Artificial Intelligence: A Modern Approach., Pearson, ISBN-10 : 0136042597, 2009.

[15] S. Guido e A. C. Mueller, Introduction to Machine Learning with Python: A Guide for Data Scientists., Sebastopol, CA: O'Reilly Media, Inc., 2016.

[16] B. Ferwerda e M. Tkalcic, «Predicting Users' Personality from Instagram Pictures: Using Visual and/or Content Features?,» in *Conference on User Modelling, Adaptation and Personalization (UMAP).*, 2018.

[17] C. J. Qian, J. D. Tang, M. A. Penza e C. M. Ferri, «Instagram Popularity Prediction via Neural Networks and Regression Analysis.,» 2019. [Online]. Available: http://cjqian.github.io/docs/instagram_paper.pdf.

[18] T. A. Poggio e F. Anselmi, «MIT Press: Visual Cortex and Deep Networks - Learning Invariant Representations.,» 2016. [Online]. Available: https://mitpress.mit.edu/books/visual-cortex-and-deep-networks.

[19] S. Jawed, U. A. Hafeez , A. S. Malik e I. Faye, «Classification of Visual and Non-visual Learners Using Electroencephalographic Alpha and Gamma Activities.,» *Frontiers in Behavioral Neuroscience.,* 2019.

[20] B. F. Timothy, T. Konkle, G. A. Alvarez e O. Aude, «Visual long-term memory has a massive storage capacity for object details.,» *Proceedings of the National Academy of Sciences of the United States of America.,* pp. 14325-14329, 2008.

[21] R. R. Hunt e J. B. Worthen, Distinctiveness and Memory., Oxford University Press, 2006.

[22] M. M. Bradley , M. K. Greenwald, M. C. Petry e P. J. Lang, «Remembering pictures: pleasure and arousal in memory.,» *Journal of Experimental Psychology: Learning, Memory and Cognition.,* vol. 18, n. 2, pp. 379-390, 1992.

[23] E. A. Phelps , «Human emotion and memory: interactions of the amygdala and hippocampal complex.,» *Current Opinion in Neurobiology.,* vol. 14, n. 2, pp. 198-202, 2005.

[24] A. Sartori, D. Culibrk, Y. Yan e N. Sebe, «Who's Afraid of Itten: Using the Art Theory of Color Combination to Analyze Emotions in Abstract Paintings.,»

Proceedings of the 23rd ACM international conference on Multimedia., pp. 311-320, 2015.

[25] A. J. Champandard, «Semantic Style Transfer and Turning Two-Bit Doodles into Fine Artworks.,» in *Computer Vision and Pattern Recognition.*, 2016.

[26] L. Standing, J. Conezio e R. N. Harber, «Perception and memory for pictures: Single-trial learning of 2500 visual stimuli.,» *Psychonomic Science.,* vol. 19, pp. 73-74, 1970.

[27] W. A. Bainbridge, «The Resiliency of Memorability: A Predictor of Memory Separate from Attention and Priming.,» *Quantitative Biology: Neurons and Cognition.,* 2017.

[28] J. Huo, «An image complexity measurement algorithm with visual memory capacity and an EEG study.,» in *SAI Computing Conference.*, 2016.

[29] M. Soleymani, «The Quest for Visual Interest.,» in *MM '15: Proceedings of the 23rd ACM international conference on Multimedia.*, 2015.

[30] F. Katsuki e C. Constantinidis , «Bottom-up and top-down attention: different processes and overlapping neural systems.,» in *The Neuroscientist.*, 2013.

[31] W. A. Bainbridge , D. D. Dilks e O. Aude, «Memorability: A stimulus-driven perceptual neural signature distinctive from memory.,» *NeuroImage.,* vol. 149, n. 1, pp. 141-152, 2017.

[32] J. Hargrave, «True Center Publishing: Integrative Works in Psychology and the Arts: Psychological Reactions to Post-Processing in Photography.,» 2013. [Online]. Available: http://truecenterpublishing.com/photopsy/postprocessing.pdf.

[33] K. Suzuki e R. Takahashi , «Effectiveness of color in picture recognition memory.,» *Japanese Psychological Research.,* vol. 39, n. 1, pp. 25-32, 1997.

[34] R. Eisenman, «Pleasing and Interesting Visual Complexity: Support for Berlyne.,» *Perceptual and Motor Skills.,* vol. 23, pp. 1167-1170, 1966.

[35] D. E. Berlyne, Conflict, arousal, and curiosity., McGraw-Hill Book Company., 1960.

[36] L. F. Barrett e M. M. Tugade, «Individual Differences in Working Memory Capacity and Dual-Process Theories of the Mind.,» *Psychological Bulletin Journal.,* vol. 130, n. 4, pp. 553-573, 2004.

[37] A. Gruszka, G. Matthews e B. Szymura, Handbook of Individual Differences in Cognition: Attention, Memory, and Executive Control., Springer, 2010.

[38] C. Jarrold e J. N. Towse, «Individual differences in working memory.,» *The Neuroscience.,* vol. 139, n. 1, pp. 39-50, 2006.

[39] P. Isola, X. Jianxiong, D. Parikh, A. Torralba e O. Aude, «What makes a photograph memorable?,» in *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 2013.

[40] A. Khosla, X. Jianxiong , P. Isola, A. Torralba e O. Aude, «Image Memorability and Visual Inception.,» *SIGGRAPH Asia 2012 Technical Briefs.,* pp. 1-4, 2012.

[41] M. Gygli, H. Grabner, H. Riemenschneider, F. Nater e L. Van Gool, «The Interestingness of Images.,» in *IEEE International Conference on Computer Vision.*, 2013.

[42] R. Halonen, S. Westman e P. Oittinen, «Naturalness and interestingness of test images for visual quality evaluation.,» *Image Quality and System Performance VIII, 78670Z.,* 2011.

[43] P. P. Aitken, «Judgments of pleasingness and interestingness as functions of visual complexity.,» *Journal of Experimental Psychology.,* vol. 103, n. 2, pp. 240-244, 1974.

[44] H. Grabner, F. Nater, M. Druey e L. Van Gool, «Visual interestingness in image sequences.,» in *Proceedings of the 21st ACM international conference on Multimedia.*, 2013.

[45] A. Forsythe, «Visual Complexity: Is That All There Is?,» in *International Conference on Engineering Psychology and Cognitive Ergonomics. Lecture Notes in Computer Science, vol 5639. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02728-4_17*, 2009.

[46] R. K. Aslanoğlu, «Color and visual complexity in abstract images.,» in *Color Research & Application.*, 2018.

[47] V. C. Müller, Philosophy and Theory of Artificial Intelligence, Springer, 2017.

[48] J. Nielsen e K. Pernice, Eyetracking Web Usability., New Riders, ISBN 0321714075, 2009.

[49] M. Kenneway, «Marketing the library: using technology to increase visibility, impact and reader engagement.,» 2007. [Online]. Available: https://serials.uksg.org/articles/abstract/10.1629/2092/.

[50] J. Wang, A. A. Nabi, G. Wang, W. Chengde e N. Tian-Tsong, «Towards Predicting the Likeability of Fashion Images.,» in *Computer Vision and Pattern Recognition.*, 2015.

[51] A. Deza e D. Parkih, «Understanding Image Virality.,» in *Conference on Computer Vision and Pattern Recognition (CVPR).*, 2015.

[52] R. C. Clark, Graphics for Learning: Proven Guidelines for Planning, Designing, and Evaluating Visuals in Training Materials., Pfeiffer, 2010.

[53] J. Sweller e P. Chandler, «Why Some Material Is Difficult to Learn?,» *Cognition and Instruction,* vol. 12, n. 3, 1994.

[54] C. L. Grady, A. R. McIntosh, M. N. Rajah e F. I. Craik, «Neural correlates of the episodic encoding of pictures and words.,» *Proceedings of the National Academy of Sciences.,* vol. 95, n. 5, 1998.

[55] F. Mengjuan, W. Jiang e W. Mao, «Creating memorable video summaries that satisfy the users intention for taking the videos.,» *Neurocomputing.,* 2018.

[56] A. Khosla, «Massachusetts Institute of Technology, cataloged from PDF version of the thesis: Predicting human behavior using visual media.,» 2017. [Online]. Available: https://dspace.mit.edu/handle/1721.1/109001.

[57] M. Khosrow-Pour, Advanced Methodologies and Technologies in Digital Marketing and Entrepreneurship., IGI Global, Disseminator of Knowledge. A volume in the Advances in Marketing, Customer Relationship Management and E-Services Book Series., 2018.

[58] T. Huang, «Computer Vision: Evolution And Promise.,» *High technology imaging science and technology.,* 1996.

[59] R. Krishna, «Computer Vision: Foundations and Applications.,» Stanford University Course., 2017. [Online]. Available: http://vision.stanford.edu/teaching/cs131_fall1718/.

[60] E. R. Kandel, «An introduction to the work of David Hubel and Torsten Wiesel.,» *The Journal of Psychology.,* 2009.

[61] D. Marr e T. Poggio, «MIT Libraries: From Understanding Computation to Understanding Neural Circuitry.,» 1976. [Online]. Available: https://dspace.mit.edu/handle/1721.1/5782.

[62] S. J. Prince, Computer Vision: Models, Learning, and Inference., Cambridge University Press., 2012.

[63] L. A. Gatys, A. S. Ecker e M. Bethge, «A Neural Algorithm of Artistic Style.,» in *Computer Vision and Pattern Recognition.*, 2015.

[64] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless e D. H. Salesin, «Image Analogies.,» in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques (SIGGRAPH)*, 2001.

[65] K.-C. Peng, T. Chen, A. Sadovnik e A. Gallagher, «A Mixed Bag of Emotions: Model, Predict, and Transfer Emotion Distributions.,» in *CVPR*, 2015.

[66] A. Khosla, W. A. Bainbridge, A. Torralba e O. Aude, «Modifying the Memorability of Face Photographs.,» in *IEEE International Conference on Computer Vision.*, 2013.

[67] A. Gupta, Machine Learning with MATLAB., MATLAB Central File Exchange., 2014.

[68] W. Wenguan e S. Jianbing , «Deep Cropping via Attention Box Prediction and Aesthetics Assessment.,» in *CVPR*, 2017.

[69] K. Hye-Rin, H. Kang e L. In-Kwon, «Image Recoloring with Valence-Arousal Emotion Model.,» Wiley Online Library: Computer Graphics:., 2016. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13018.

[70] D. Furness, «Digital Trends: Incredible AI app can 'repaint' photos to look like it was composed by famous artists.,» 2016. [Online]. Available: https://bit.ly/3niS1gl.

[71] R. S. Sutton e A. G. Barto, Reinforcement Learning: An Introduction., The MIT Press., 2018.

[72] D. Ulyanov, A. Vedaldi e V. Lempitsky, «Improved Texture Networks: Maximizing Quality and Diversity in Feed-forward Stylization and Texture Synthesis.,» in *Computer Vision and Pattern Recognition.*, 2017.

[73] X. Huang e S. Belongie, «Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization.,» in *Computer Vision and Pattern Recognition.*, 2017.

[74] F. Luan, S. Paris, E. Shechtman e K. Bala, «Deep Photo Style Transfer.,» in *Computer Vision and Pattern Recognition.*, 2017.

[75] H. Zhang e K. Dana, «Multi-style Generative Network for Real-time Transfer.,» in *CVPR*, 2017.

[76] T. Sanocki e N. Sulman, «Color relations increase the capacity of visual short-term memory.,» *Perception.,* vol. 40, n. 6, pp. 635-648, 2011.

[77] L. A. Gatys, A. S. Ecker e M. Bethge, «Image Style Transfer Using Convolutional Neural Networks.,» in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[78] M. Ruder, A. Dosovitskiy e T. Brox, «Artistic style transfer for videos and spherical images.,» in *Computer Vision and Pattern Recognition.*, 2018.

[79] L. Chuan e M. Wand, «Precomputed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks.,» in *Computer Vision and Pattern Recognition.*, 2016.

[80] A. Krizhevsky, I. Sutskever e G. E. Hinton, «ImageNet Classification with Deep Convolutional Neural Networks.,» *Advances in Neural Information Processing Systems 25 (NIPS).,* 2012.

[81] K. J. Holmes, C. Alcat e S. F. Lourenco, «Is Emotional Magnitude Spatialized? A Further Investigation.,» *Cognitive Science.,* vol. 43, n. 4, 2019.

[82] B. Zohu, A. Lapedriza, J. Xiao, A. Torralba e O. Aude, «Learning Deep Features for Scene Recognition using Places Database.,» *Advances in Neural Information Processing Systems 27 (NIPS).,* 2014.

[83] D. Ulyanov, V. Lebedev, A. Vedaldi e V. Lempitsky, «Texture Networks: Feed-forward Synthesis of Textures and Stylized Images.,» in *Computer Vision and Pattern Recognition.*, 2016.

[84] A. Sartori, V. Yanulevskaya, A. A. A. Salah e J. R. R. Uijlings, «Affective Analysis of Professional and Amateur Abstract Paintings Using Statistical Analysis and Art Theory.,» *The ACM Transactions on Interactive Intelligent Systems.,* vol. 5, n. 2, 2015.

[85] J. Machajdik e A. Hanbury, «Affective image classification using features inspired by psychology and art theory.,» *Proceedings of the 18th ACM international conference on Multimedia.,* pp. 83-92, 2010.

[86] D. R. Lide, Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables., A Century of Excellence in Measurements, Standards and Technology., 2018.

[87] A. Khosla, J. Xiao, A. Torralba e O. Aude, «Memorability of Image Regions.,» in *Advances in Neural Information Processing Systems (NIPS).*, 2012.

[88] K. Simonyan e A. Zisserman, «Very Deep Convolutional Networks for Large-Scale Image Recognition.,» in *Computer Vision and Pattern Recognition.*, 2014.

[89] J. S. Warm, R. Parasuraman e G. Matthews, «Vigilance Requires Hard Mental Work and Is Stressful.,» *Human Factors The Journal of the Human Factors and Ergonomics Society.,* vol. 50, n. 3, pp. 433-41, 2008.

[90] D. E. Berlyne, «Complexity and incongruity variables as determinants of exploratory choice and evaluative ratings.,» *Canadian Journal of Psychology/Revue canadienne de psychologie.,* vol. 17, n. 3, pp. 274-290, 1963.

[91] A. Siarohin, G. Zen, X. Alameda-Pineda, E. Ricci e N. Sebe, «Increasing image memorability with neural style transfer.,» *ACM Transactions on Multimedia Computing, Communications, and Applications.,* vol. 15, n. 2, 2019.

[92] S. Shekhar, S. P. M. Angara, M. Kedia, D. Singal e A. S. Shetty, «Techniques for enhancing content memorability of user generated video content.,» 2015. [Online]. Available: https://patents.google.com/patent/US10380428B2/en.

[93] R. Klette, Concise Computer Vision: An Introduction into Theory and Algorithms., Springer, 2014.

[94] L. He, H. Qi e R. Zaretzki, «Image color transfer to evoke different emotions based on color combinations.,» *Computer Science: Signal, Image and Video Processing.,* 2015.

[95] R. A. Afsheen e A. Mohsen, «Emotional Filters: Automatic Image Transformation for Inducing Affect.,» in *Computer Vision and Pattern Recognition.*, 2017.

[96] L. Wang, N. Xiang, Y. Xiaosong e Z. Jianjun, «Fast photographic style transfer based on convolutional neural networks.,» *Proceedings of Computer Graphics International.,* pp. 67-68, 2018.

[97] L. Sheng, Z. Lin, S. Jing e W. Xiaogang, «Avatar-Net: Multi-scale Zero-shot Style Transfer by Feature Decoration.,» in *Computer Vision and Pattern Recognition.*, 2018.

[98] B. Zhou, À. Lapedriza, A. Khosla e O. Aude, «Places: A 10 Million Image Database for Scene Recognition.,» in *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 2017.

[99] A. R. Smith, «ResearchGate: Image Compositing Fundamentals.,» 2000. [Online]. Available: https://www.researchgate.net/publication/2350869_Image_Compositing_Fundamentals.

[100] R. Hall e D. Greenberg, «A Testbed for Realistic Image Synthesis.,» *IEEE Computer Graphics and Applications.,* vol. 3, pp. 10-20, 1983.

[101] L. Hsiang-Ming , L. Ching-Chi e W. Cou-Chen , «Brand image strategy affects brand equity after M&A.,» *European Journal of Marketing.,* vol. 45, n. 7, pp. 1091-1111, 2011.

[102] B. K. Wiederhold, «Instagram: Becoming a Worldwide Problem?,» *Cyberpsychology, Behavior, and Social Networking.,* vol. 22, n. 9, pp. 567-568, 2019.

Образац - План третмана података

Овај Образац чини саставни део докторске дисертације, односно докторског уметничког пројекта који се брани на Универзитету у Новом Саду. Попуњен Образац укоричити иза текста докторске дисертације, односно докторског уметничког пројекта.

План третмана података

Назив пројекта/истраживања

Аутоматско повећање памтљивости слика

Назив институције/институција у оквиру којих се спроводи истраживање

а) Универзитет у Новом Саду, Факултет техничких наука
б) Универзитет у Тренту, Департман за информационо инжењерство и рачунарске науке
в) Стеленбош Универзитет и Алкала Универзитет – студент у посети

Назив програма у оквиру ког се реализује истраживање

Споразум о међународном двојном докторату (*Co-tutelle de thèse*) између Универзитета у Новом Саду, Факултета техничких наука (Индустријско инжењерство и инжењерски менаџмент) и Универзитета у Тренту.
Тема: Аутоматско повећање памтљивости слика.

1. Опис података

1.1 Врста студије

Укратко описати тип студије у оквиру које се подаци прикупљају
Коришћење техника машинског учења за постизање вештачке интелигенције подразумева примену алгоритама претходно истренираних подацима. Рачунари овом приликом користе велике количине података како би извршили одређени задатак и изградили модел са циљем предвиђања одређених појава или доношења одлука. Током реализације истраживања приступило се квантитативној и квалитативној евалуацији предложеног модела за аутоматско повећање памтљивости слика. Емпиријско истраживање посматраних и мерљивих карактеристика слике служили су за проверу теоријских оквира, предикцију резултата и утврђивање односа између варијабли коришћењем статистичких анализа. Циљ прикупљања података био је демонстрација ефективности предложеног приступа посредством експерименталне студије на јавно доступном скупу података ЛаМем (енг. *Large-Scale Image Memorability – LaMem*) који представља највећу постојећу базу анотираних слика за процену памтљивости, као и применом корисничке студије за чије потребе је реализована игра меморије. Осим овога, за потребе истраживања коришћен

је и сет апстрактних уметничких слика тзв. *DevianaArt*, због процене да је апстрактна уметност која се углавном ослања на комбинације текстуре и боје, најпогоднија за ово истраживање.

1.2 Врсте података
а) квантитативни – памтљивост слике је сама по себи објективна и квантитативна мера и независна је од посматрача, а може се и рачунарски предвидети. Из тог разлога је највећи део студије посвећен оваквој врсти анализе података.
б) квалитативни – за потребе истраживања и евалуације модела, реализована је и корисничка студија и конструисана мала игра меморије током које су испитаници, учесници у истраживању, допринели додатној валидацији.

1.3. Начин прикупљања података
а) анкете, упитници, тестови
б) клиничке процене, медицински записи, електронски здравствени записи
в) генотипови: навести врсту ________________________________
г) административни подаци: навести врсту _______________________
д) узорци ткива: навести врсту_________________________________
ђ) снимци, фотографије: навести врсту - **фотографије**
е) текст, навести врсту ______________________________________
ж) мапа, навести врсту ______________________________________
з) остало: описати ____

1.3 Формат података, употребљене скале, количина података
ЛаМем база, укупно 58.741 фотографија коришћено (http://memorability.csail.mit.edu/). Такође, 500 апстрактних слика (http://disi.unitn.it/~sartori/datasets/deviantart-dataset/). Током експерименталне фазе, коришћено је 45000 слика за тренинг, 10000 слика за тестирање и 3741 слика за валидацију овог истераживања.

1.3.1 Употребљени софтвер и формат датотеке:
a) Excel фајл, датотека __________________
b) SPSS фајл, датотека __________________
c) PDF фајл, датотека ___________________
d) Текст фајл, датотека __________________
e) JPG фајл, датотека – сет података сачињен од фајлова у формату *.jpg*
f) Остало, датотека ____________________

1.3.2. Број записа (код квантитативних података)
а) број варијабли – једна
б) број мерења (испитаника, процена, снимака и сл.):

- за тренинг је изведено 70.000 итерација стохастичког градијентног спуштања (енг. *stochastic gradient descent*) са моментумом 0.9, брзином учења (енг. *learning rate*) 10^{-3} и величином серије (енг. *batch size*) 256;
- у корисничкој студији (игри меморије) и процесу валидације, учествовало је укупно 200 испитаника.

1.3.3. Поновљена мерења
а) да
б) не

Уколико је одговор да, одговорити на следећа питања:
а) временски размак између поновљених мера је ______________________________
б) варијабле које се више пута мере односе се на ________________________________
в) нове верзије фајлова који садрже поновљена мерења су именоване као ____________

Напомене: ____

Да ли формати и софтвер омогућавају дељење и дугорочну валидност података?
а) Да
б) Не
Ако је одговор не, образложити ____
____

2. Прикупљање података

2.1 Методологија за прикупљање/генерисање података
За потребе истраживања примењени су експерименти на јавно доступним сетовима података претходно описаним под тачком 1.3. Дисертација се ослања на актуелно стање из домена памтљивости слика, полазећи од прикупљених скупова података слика посебно дизајнираних за проучавање овог својства фотографије, укључујући и технике коришћене за анотацију коришћених сетова података скоровима памтљивости. ЛаМем сет података представља колекцију насталу прикупљањем неколико постојећих, различитих сетова података, укључујућу сето података афективних слика, сачињеног од уметничких и апстрактних слик, коришћеног у великом броју истраживања емоција. Апстрактне слике из сета ДевиантАрт прикупљене су са онлајн веб страница и социјалних мрежа намењених дељењу уметничких слика које су креирали сами корисници. Представља једну од највећих интернет уметничких заједница (са око 360 милиона уметничких дела и преко 35 милиона регистрованих корисника, међу којима има и аматера и професионалних уметника). Све слике организоване су у категорије, што поједностављује преузимање само жељеног стила. За потребе реализације ове студије коришћене су апстрактне слике из категорије: Традиционално – Уметност – Слике – Апстрактна уметност.

2.1.1. У оквиру ког истраживачког нацрта су подаци прикупљени?
а) експеримент, навести тип – **експерименти у рачунарском виду (енг. *Computer Vision*)**
б) корелационо истраживање, навести тип ____
ц) анализа текста, навести тип ____
д) остало, навести шта – Дескриптивна анализа података

2.1.2 Навести врсте мерних инструмената или стандарде података специфичних за одређену научну дисциплину (ако постоје).

У процесу евалуације резултата коришћени су Пирсонов коефицијент корелације (енг. *Pearson's correlation coefficient*), Спирманов коефицијент корелације (енг. *Spearman's rank correlation coefficient*), Средња грешка квадрата (енг. *Mean squared error*), Tačnost вредности величине (енг. *Accuracy*).

Осим квантитативне, спроведена је и квалитативна евалуација и креирана игра меморије, праћењем протокола из претходно објављених радова.

2.2 Квалитет података и стандарди

Пречишћавање и трансформација података пре уласка у процес анализе.

2.2.1. Третман недостајућих података

а) Да ли матрица садржи недостајуће податке? Да **Не**

Ако је одговор да, одговорити на следећа питања:

а) Колики је број недостајућих података? ___________________________

б) Да ли се кориснику матрице препоручује замена недостајућих података? Да Не

в) Ако је одговор да, навести сугестије за третман замене недостајућих података

____

2.2.2. На који начин је контролисан квалитет података? Описати

Искази о томе да су подаци, односно сетови података (фотографија), коришћени за потребе овог истраживања доследни, тачни, потпуни и веродостојни пронађени су у самом опису скупова података, као и у пруженој експерименталној валидаацији. За креирање скупа података знатно разноврснијег од свих до тада постојећих, коришћени су најразличитији скупови података и истраживана је корелација између атрибута везаних за сваки појединачни скуп и памтљивости.

2.2.3. На који начин је извршена контрола уноса података у матрицу?

Коришћени подаци припадају претходно описаним јавно доступним скуповима података. Пре уноса податакаа у моделе машинског учења током истраживања, извршена је провера увидом у иницијалну базу података.

3. Третман података и пратећа документација

3.1. Третман и чување података

3.1.1. Подаци ће бити депоновани у НаРДзС репозиторијуму (Национални Репозиторијум Дисертација у Србији).

3.1.2. URL адреса https://www.cris.uns.ac.rs/searchDissertations.jsf

3.1.3. DOI

____

3.1.4. Да ли ће подаци бити у отвореном приступу?
а) Да
б) Да, али после ембарга који ће трајати до ______________________________
в) Не

Ако је одговор не, навести разлог ______________________________

3.1.5. Подаци неће бити депоновани у репозиторијум, али ће бити чувани.
Образложење

____

3.2 Метаподаци и документација података
3.2.1. Који стандард за метаподатке ће бити примењен?
Стандард који примењује Репозиторијум Универзитета у Новом Саду.
3.2.1. Навести метаподатке на основу којих су подаци депоновани у репозиторијум.
Цвета Мајтановић (2021): Аутоматско повећање памтљивости слика

____

Ако је потребно, навести методе које се користе за преузимање података, аналитичке и процедуралне информације, њихово кодирање, детаљне описе варијабли, записа итд.

____

____

3.3 Стратегија и стандарди за чување података
3.3.1. До ког периода ће подаци бити чувани у репозиторијуму? Нема ограничења.
3.3.2. Да ли ће подаци бити депоновани под шифром? Да **Не**
3.3.3. Да ли ће шифра бити доступна одређеном кругу истраживача? Да **Не**
3.3.4. Да ли се подаци морају уклонити из отвореног приступа после извесног времена?
Да **Не**
Образложити

____

____

4. Безбедност података и заштита поверљивих информација

Овај одељак МОРА бити попуњен ако ваши подаци укључују личне податке који се односе на учеснике у истраживању. За друга истраживања треба такође размотрити заштиту и сигурност података.

4.1 Формални стандарди за сигурност информација/података
Истраживачи који спроводе испитивања с људима морају да се придржавају Закона о заштити података о личности (*https://www.paragraf.rs/propisi/zakon_o_zastiti_podataka_o_licnosti.html*) и одговарајућег институционалног кодекса о академском интегритету.

4.1.2. Да ли је истраживање одобрено од стране етичке комисије? Да Не
Ако је одговор Да, навести датум и назив етичке комисије која је одобрила истраживање
Истраживање је спроведено у складу са прописима за истраживања овог типа на Универзитету у Тренту. За потребе истраживања нису прикупљане никакве личне информације од учесика.

4.1.2. Да ли подаци укључују личне податке учесника у истраживању? Да **Не**
Ако је одговор да, наведите на који начин сте осигурали поверљивост и сигурност информација везаних за испитанике:
а) Подаци нису у отвореном приступу
б) Подаци су анонимизирани
ц) Остало, навести шта

5. Доступност података

5.1. Подаци ће бити
а) јавно доступни
б) доступни само уском кругу истраживача у одређеној научној области
ц) затворени

Ако су подаци доступни само уском кругу истраживача, навести под којим условима могу да их користе:

Ако су подаци доступни само уском кругу истраживача, навести на који начин могу приступити подацима:

5.4. Навести лиценцу под којом ће прикупљени подаци бити архивирани.
Ауторство – некомерцијално – без прераде

6. Улоге и одговорност

6.1. Навести име и презиме и мејл адресу власника (аутора) података
Цвета Мајтановић, cveta.majtanovic@gmail.com

6.2. Навести име и презиме и мејл адресу особе која одржава матрицу с подацима
Цвета Мајтановић, cveta.majtanovic@gmail.com

6.3. Навести име и презиме и мејл адресу особе која омогућује приступ подацима другим истраживачима
Цвета Мајтановић, cveta.majtanovic@gmail.com