Bogdan Ionescu
Wilma A. Bainbridge
Naila Murray *Editors*

Human Perception of Visual Information

Psychological and Computational Perspectives

Springer

Human Perception of Visual Information

Bogdan Ionescu • Wilma A. Bainbridge
Naila Murray
Editors

Human Perception of Visual Information

Psychological and Computational Perspectives

Springer

Editors
Bogdan Ionescu
Politehnica University of Bucharest
Bucharest, Romania

Wilma A. Bainbridge
University of Chicago
Chicago, IL, USA

Naila Murray
NAVER Labs Europe
Meylan, France

ISBN 978-3-030-81464-9 ISBN 978-3-030-81465-6 (eBook)
https://doi.org/10.1007/978-3-030-81465-6

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

There is one thing the photograph must contain, the humanity of the moment.

—Robert Frank

Computational models of objective visual properties such as semantic content and geometric relationships have made significant breakthroughs using the latest achievements in machine learning and large-scale data collection. There has also been limited but important work exploiting these breakthroughs to improve computational modelling of subjective visual properties such as interestingness, affective values and emotions, aesthetic values, memorability, novelty, complexity, visual composition and stylistic attributes, and creativity. Researchers that apply machine learning to model these subjective properties are often motivated by the wide range of potential applications of such models, including for content retrieval and search, storytelling, targeted advertising, education and learning, and content filtering. The performance of such machine learning-based models leaves significant room for improvement and indicates a need for fundamental breakthroughs in our approach to understanding such highly complex phenomena.

Largely in parallel to these efforts in the machine learning community, recent years have witnessed important advancements in our understanding of the psychological underpinnings of these same subjective properties of visual stimuli. Early focuses in the vision sciences were on the processing of simple visual features like orientations, eccentricities, and edges. However, utilizing new neuroimaging techniques such as functional magnetic resonance imaging, breakthroughs through the 1990s and 2000s uncovered specialized processing in the brain for high-level visual information, such as image categories (e.g., faces, scenes, tools, objects) and more complex image properties (e.g., real-world object size, emotions, aesthetics). Recent work in the last decade has leveraged machine learning techniques to allow researchers to probe the specific content of visual representations in the brain. In parallel, the widespread advent of the Internet has allowed for large-scale crowd-sourced experiments, allowing psychologists to go beyond small samples with limited, controlled stimulus sets to study images at a large scale. With the combination of these advancements, psychology is now able to take a fresh look at

age-old questions like what we find interesting, what we find beautiful, what drives our emotions, how we perceive spaces, or what we remember.

The field of machine learning, and Artificial Intelligence more broadly, enjoys a long tradition of seeking inspiration from investigations into the psychology and neuroscience of human and non-human intelligence. For example, deep learning neural networks in Computer Vision were originally inspired by the architecture of the human visual system, with its many layers of neurons thought to apply filters at each stage. Psychology and neuroscience also rely heavily on developments from Artificial Intelligence, both for parsing down the Big Data collected from the brain and behavior, as well as for understanding the underlying mechanisms. For example, now, object classification deep neural networks such as VGG-16 are frequently used as stand-ins for the human visual system to predict behavior or even activity in the brain. Given the progress made in machine learning and psychology towards more successfully modelling subjective visual properties, we believe that the time is ripe to explore how these advances can be mutually enriching and lead to further progress.

To that end, this book showcases complementary perspectives from psychology and machine learning on high-level perception of images and videos. It is an interdisciplinary volume that brings together experts from psychology and machine learning in an attempt to bring these two, at a first glance, different fields, into conversation, while at the same time providing an overview of the state of the art in both fields. The book contains 10 chapters arranged in 5 pairs, with each pair describing state-of-the-art psychological and computational approaches to describing and modelling a specific subjective perceptual phenomenon.

In Chap. 1, Lauer and Võ review recent studies that use diverse methodologies like psychophysics, eye tracking, and neurophysiology to help better capture human efficiency in real-world scene and object perception. The chapter focuses in particular on which contextual information humans take advantage of most and when. Further, they explore how these findings could be useful in advancing computer vision and how computer vision could mutually further understanding of human visual perception. In Chap. 2, Constantin et al. consider the related phenomenon of interestingness prediction from a computational point of view and present an overview of traditional fusion mechanisms, such as statistical fusion, weighted approaches, boosting, random forests, and randomized trees. They also include an investigation of a novel, deep learning-based system fusion method for enhancing performance of interestingness prediction systems.

In Chap. 3, Bradley et al. review recent research related to photographic images that depict affectively engaging events, with the goal of assessing the extent to which specific pictures reliably engage emotional reactions across individuals. In particular, they provide preliminary analyses that encourage future investigations aimed at constructing normative biological image databases that, in addition to evaluative reports, provide estimates of emotional reactions in the body and brain for use in studies of emotion and emotional dysfunction. On the computational side, in Chap. 4, Zhao et al. introduce image emotion analysis from a computational perspective with a focus on summarizing recent advances. They revisit key computational

problems with emotion analysis and present in detail aspects such as emotion feature extraction, supervised classifier learning, and domain adaptation. Their discussion concludes with the presentation of the relevant datasets for evaluation and the identification of open research directions.

In Chap. 5, Chamberlain sets out the history of empirical aesthetics in cognitive science and the state of the research field at present. The chapter outlines recent work on inter-observer agreement in aesthetic preference before presenting empirical work that argues the importance of objective (characteristics of stimuli) and subjective (characteristics of context) factors in shaping aesthetic preference. Valenzise et al. explore machine learning approaches to modelling computational image aesthetics, in Chap. 6. They overview the several interpretations that aesthetics have received over time and introduce a taxonomy of aesthetics. They discuss computational advances in aesthetics prediction, from early methods to deep neural networks, and overview the most popular image datasets. Open challenges are identified and discussed, including dealing with the intrinsic subjectivity of aesthetic scores and providing explainable aesthetic predictions.

Bainbridge, in Chap. 7, draws from neuroimaging and other research to describe our current state-of-the-art understanding of memorability of visual information. Such research has revealed that the brain is sensitive to memorability both rapidly and automatically during late perception. These strong consistencies in memory across people may reflect the broad organizational principles of our sensory environment and may reveal how the brain prioritizes information before encoding items into memory. In Chap. 8, Bylinskii et al. examine the notion of memorability with a computational lens, detailing the state-of-the-art algorithms that accurately predict image memorability relative to human behavioral data, using image features at different scales from raw pixels to semantic labels. Beyond prediction, they show how recent Artificial Intelligence approaches can be used to create and modify visual memorability, and preview the computational applications that memorability can power, from filtering visual streams to enhancing augmented reality interfaces.

In Chap. 9, Akcelik et al. review recent research that aims to quantify visual characteristics and design qualities of built environments, in order to relate more abstract aspects of an urban space to quantifiable design features. Uncovering these relationships may provide the opportunity to establish a causal relationship between design features and psychological feelings such as walkability, preference, visual complexity, and disorder. Lastly, in Chap. 10, Medina Ríos et al. review research that uses machine learning approaches to study how people perceive urban environments according to subjective dimensions like beauty and danger. Then, with a specific focus on Global South cities, they present a study on perception of urban scenes by people and machines. They use their findings from this study to discuss implications for the design of systems that use crowd-sourced subjective labels for machine learning and inference on urban environments.

We have edited this book to appeal to undergraduate and graduate students, academic and industrial researchers, and practitioners who are broadly interested in cognitive underpinnings of subjective visual experiences, as well as computational approaches to modelling and predicting them. The authors of this book provide

overviews of the current state of the art in their respective fields of study; therefore, chapters are largely accessible to researchers who may not be familiar with either prevailing computational, and particularly machine learning, practice, or with research practice in cognitive science. As such, we believe that researchers from both worlds will have much to learn from these chapters.

We are indebted to all the authors for their contributions, and hope that readers of this book will enjoy reading the fruits of their hard work as much as we have. Finally, we thank our editor, Springer, who gave us the opportunity to bring this project to life.

Bucharest, Romania — Bogdan Ionescu

Chicago, IL, USA — Wilma A. Bainbridge

Meylan, France — Naila Murray

Contents