

Accepted Manuscript

From Pixels to Sentiment: Fine-tuning CNNs for Visual Sentiment Prediction

Víctor Campos, Brendan Jou, Xavier Giró-i-Nieto

PII: S0262-8856(17)30035-5
DOI: doi:[10.1016/j.imavis.2017.01.011](https://doi.org/10.1016/j.imavis.2017.01.011)
Reference: IMAVIS 3596

To appear in: *Image and Vision Computing*

Received date: 9 April 2016
Revised date: 19 January 2017
Accepted date: 26 January 2017



Please cite this article as: Víctor Campos, Brendan Jou, Xavier Giró-i-Nieto, From Pixels to Sentiment: Fine-tuning CNNs for Visual Sentiment Prediction, *Image and Vision Computing* (2017), doi:[10.1016/j.imavis.2017.01.011](https://doi.org/10.1016/j.imavis.2017.01.011)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

From Pixels to Sentiment: Fine-tuning CNNs for Visual Sentiment Prediction

Víctor Campos^a, Brendan Jou^b, Xavier Giró-i-Nieto^c

^aBarcelona Supercomputing Center (BSC), Barcelona, Catalonia/Spain

^bColumbia University, New York, NY USA

^cUniversitat Politècnica de Catalunya (UPC), Barcelona, Catalonia/Spain

Abstract

Visual multimedia have become an inseparable part of our digital social lives, and they often capture moments tied with deep affections. Automated visual sentiment analysis tools can provide a means of extracting the rich feelings and latent dispositions embedded in these media. In this work, we explore how Convolutional Neural Networks (CNNs), a now de facto computational machine learning tool particularly in the area of Computer Vision, can be specifically applied to the task of visual sentiment prediction. We accomplish this through fine-tuning experiments using a state-of-the-art CNN and via rigorous architecture analysis, we present several modifications that lead to accuracy improvements over prior art on a dataset of images from a popular social media platform. We additionally present visualizations of local patterns that the network learned to associate with image sentiment for insight into how visual positivity (or negativity) is perceived by the model.

Keywords: Sentiment, Convolutional Neural Networks, Social Multimedia, Fine-tuning Strategies

1. Introduction

The sheer throughput of user-generated multimedia content uploaded to social networks every day has experienced tremendous growth in the last several years. These social networks often serve as platforms for their users to express feelings and opinions. And visual multimedia, in particular, has become a natural and rich form to communicate emotions and sentiments in a host of these digital media platforms.

Affective Computing [1] is lately drawing increased attention by multiple research disciplines. This increased interest may be attributed to recent successes in areas like emotional understanding of viewer responses to advertisements using facial expressions [2] and monitoring of emotional patterns to help patients suffering from mental health disorder [3]. Given the complexity of the task, visual understanding for emotion and sentiment detection has lagged behind other Computer Vision tasks, e.g., in general object recognition.

Emotion and *sentiment* are closely connected entities. Emotion is usually defined as high intensity, but relatively brief experience, onset by a stimuli [4, 5],

whereas sentiment refers to an attitude, disposition or opinion towards a certain topic [6] and usually implies a longer-lived phenomena than that in emotion. Throughout this work we represent sentiment values as a polarity that can be either *positive* or *negative*, although some works also consider the *neutral* class or even a finer scale that accounts for different strengths [7]. Since the data used in our experiments is annotated using crowd-sourcing, we believe that binary binning was helpful to force the annotators to decide between either polarities rather than tend toward a neutral rating.

The state-of-the-art in classical Computer Vision tasks have recently undergone rapid transformations thanks to the re-popularization of Convolutional Neural Networks (CNNs) [8, 9]. This led us to also explore such architectures for visual sentiment prediction where we seek to recognize the sentiment that an image would provoke to a human viewer. Given the challenge of collecting large-scale datasets with reliable sentiment annotations, our efforts focus on understanding domain-transferred CNNs for visual sentiment prediction by analyzing the performance of a state-of-the-art architecture fine-tuned for this task.

In this paper, we extend our previous work in [10], where we empirically studied the suitability of domain transferred CNNs for visual sentiment prediction. The

Email addresses: victor.campos@bsc.es (Víctor Campos),
bjou@caa.columbia.edu (Brendan Jou), xavier.giro@upc.edu
(Xavier Giró-i-Nieto)

Download English Version:

<https://daneshyari.com/en/article/4968955>

Download Persian Version:

<https://daneshyari.com/article/4968955>

[Daneshyari.com](https://daneshyari.com)