Deep Learning Approaches on the Recognition of Affective Properties of Images

Takahisa Yamamoto

Abstract

Visual stimuli in images evoke various affective properties, including affections, impressions, feelings, and so forth. The ability to understand and recognize the affective properties of images is crucial in many research areas, such as affective computing, impression management, and kansei engineering, and has many useful applications.

In regard to affective computing, the ability of recognizing emotions is important for achieving natural interactions and intelligent communications between humans and machines, since machines without emotions cannot understand our feelings.

In regard to impression management, fashion is an important means of eliciting positive and negative impressions of persons. Having a means of objectively checking fashion style to determine if it is suitable for certain situations and places would be very useful. Moreover, with the increasing popularity of social networking services (SNSs), commercial applications suggesting impression improvements in uploading pictures are considered to be important.

Feelings are important aspects of kansei engineering, and the ability to recognize design and emotional elements is a requirement for creating attractive products reflecting consumer's feelings and desires.

This thesis deals with problems of recognizing the affective properties of images by focusing on creating affective features. Our framework for these recognition problems is to create effective features from objects and components with affective semantics based on our ideas (hypotheses) in addition to visual features from a whole input image. We treat image emotion recognition and fashion style recognition as recognition problems due to their importance in research, business, and applications. Our goal is to improve classification performance and to pave the way for new business applications. These recognition problems are still difficult than conventional problems. Several methods have been proposed, but their accuracy is still unsatisfactory.

One of the difficulties in classifying image emotions is that sentiments are evoked by different types of information, i.e., visual and semantic information. To tackle this issue, the algorithm described in this thesis uses three different feature vectors: one is a visual feature vector obtained by a convolutional neural network (CNN) that retrieves the sentiment from a whole input image. The other two semantic feature vectors are based on recognized objects and reflect the idea that several objects have emotional semantics and represent the idea that objects belonging to the same category or similar objects are related to the same emotion. Extensive evaluation experiments conducted on emotional image datasets showed that our method achieved the best accuracy except for one dataset against existing methods. We analyze the contributions of the semantic features for improving the accuracy and discuss new business applications.

One of the difficulties in classifying fashion styles is how to extract fine-grained features. Our idea is that fashion styles are represented by various components such as textures and colors of clothes, hairstyle, and so forth. The existing methods of fashion style recognition do not take various components into consideration. We find the importance of components and propose a method that separates an input image into important components for fashion style classification and extracts features from each component by using dedicated CNNs for each component in addition to a feature from a whole human area. The feature created by this algorithm has fine-grained information and is effective at distinguishing similar fashion styles. Experimental results showed that our method achieved the best accuracy in all evaluation datasets. We analyze the contributions of the components for improving the accuracy and discuss new business applications.

The structure of our thesis

Chapter 1: Introduction

This chapter describes background, motivation, affective computing, and focus and contributions of our thesis as follows.

1). Backgrounds are described from affective computing, impression management, and kansei engineering.

2). Motivations for image emotion recognition and fashion style recognition as the specific recognition problems on affective properties of images are described from their importance in research, business, and wide applications.

3). Affective computing is not main focus of this thesis, but we consider it is valuable to describe it in terms of different research areas.

4). The focus and contributions of this thesis are described. The focal point for both recognition problems is how to create affective features from recognition results of objects or components with affective semantics based on our ideas. The contributions are to propose a novel method for improving classification accuracy and to find new facts that are useful for business applications.

Chapter 2: Related Work

This chapter describes the related works on image emotion recognition and fashion style recognition.

The structure of this chapter is as follows.

1). Related works using computer vision technologies for sentiment analysis such as emotional image retrieval, emotional attention maps, are described.

2). Related works directly related to image emotion recognition are described.

3). Datasets related to emotion recognition from images are described.

4). Related works using computer vision technologies for fashion area such as searching technology for fashion items, generating fashion images, estimating fashionability, and semantic segmentation for human parsing, are described.

5). Related works directly related to fashion style recognition are described.

6). Fashion related datasets such as on semantic segmentation and fashion styles are described in detail.

Chapter 3: Image Emotion Recognition

This chapter introduces our approach to recognizing emotion evoked by an input image and indicates the usefulness by evaluation experiments.

The structure of this chapter is as follows.

1). The background and motivation of this work is introduced.

2). Our method using visual and semantic features is explained. The semantic features are created on the basis of our hypotheses about image emotion recognition. The methods of extracting a visual feature and semantic features are described.

3). Evaluation experiments using emotional datasets are conducted on our method and other methods to verify its effectiveness and usefulness. The analysis of the results shows that semantic features improve the classification accuracy. The experimental conditions and implementation information are also described.

4). Supplemental experiments using datasets having the number of annotators as meta-data are conducted. The experiments show that classification accuracy improves in accordance with the increase in agreement of the annotators.

5). Our method, experimental results, and business applications are summarized.

Chapter 4: Fashion Style Recognition

This chapter describes our approach to recognizing a fashion style depicted in an input image and shows the effectiveness through evaluation experiments.

The structure of this chapter is as follows.

1). The background and motivation of this work is introduced.

2). Our method of using multiple CNNs is explained. The multiple CNNs reflect our ideas for fashion style recognition. The methods how to preprocess, including semantic segmentation for human parsing and how to create each component from an input image are described. Our multiple CNNs extract feature vectors from each component image.

3). Evaluation experiments using fashion style datasets verify the effectiveness and usefulness of our method in comparison with other methods. The experimental conditions and implementation information are also described.

4). The effectiveness of component at improving recognition accuracy is discussed. We confirm that fashion styles can be represented by various components through experiments using only clothes image, clothes with head image, clothes and posture images, and so forth.

5). Supplemental experiments using all considerable components are conducted and analyzed.

6). Our method and experimental results are summarized. We also consider applications of our method to business.

Chapter 5: Conclusion

This chapter summarizes our method and describes our contributions in terms of affective properties recognition from images. It also describes future work.