天津
VALSE
2022

Affective Image Content Analysis: Two Decades Review and New Perspectives

Email: exped1230@exped.com
Website: cv.nankai.edu.cn

Sicheng Zhao[1], Xingxu Yao[2], Jufeng Yang[2], Guoli Jia[2], Guiguang Ding[1], Tat-Seng Chua[3], Bjorn W. Schuller[4], Kurt Keutzer[5]

[1]Tsinghua University, [2]Nankai University, [3]National University of Singapore, [4]Imperial College London, [5]University of California, Berkeley

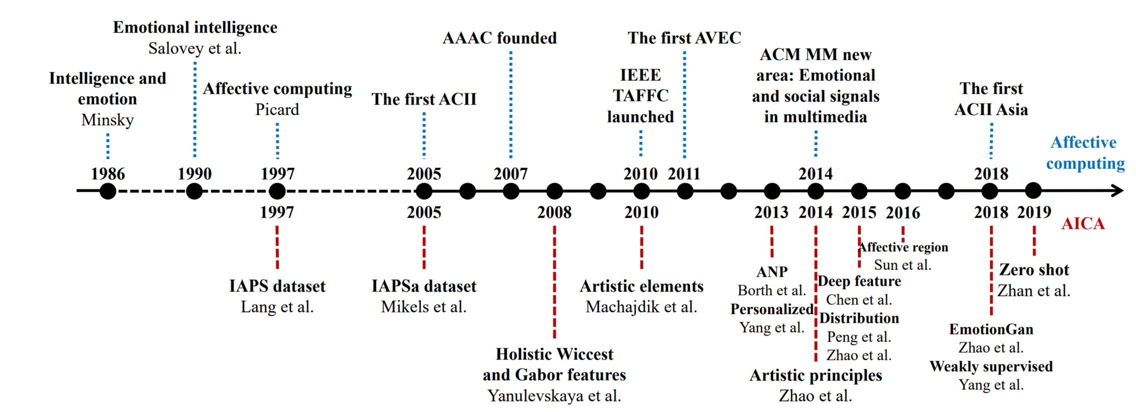
This paper has been accepted by TPAMI 2021

Introduction

Affective Image Content Analysis(AICA) aims for analyzing **viewer's feeling** after seeing the image.

➢ Emotion models:
- Categorical emotion states (CES)
- Dimensional emotion space (DES)

➢ Main contributions:
- We summarize the available **datasets** and representative **features** in the past two decades.
- we share some **potential research directions** of AICA in the future.



Brief history of AICA

Features

Features can be simply divided into hand-crafted features and deep features. One of the goal of the features is to bridge the affective gap, which is a main challenge of AICA.

➢ Hand-crafted features:
- **Low-level features**, such as color, texture.
- **Mid-level features**, such as Principles-of-art, Sentribute.
- **High-level features**, such as Sentibank, expressions.

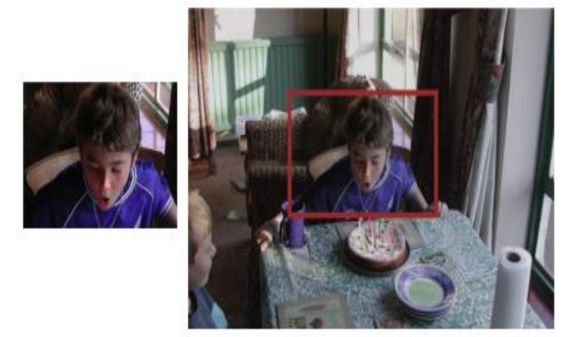
Feature	Ref	Level	Short description	# Feat
WLDLV	[89]	low	orientation and length information of lines	12
EFS	[90]	low	luminance-warm-cool fuzzy histogram, saturation-warm-cool fuzzy histogram, luminance contrast	10, 7, 2
Eleven Groups	[91]	low	shape, edge, texture, polynomial, image statistics	691
LOW_C	[7]	low	Gist, HOG2x2, self-similarity and geometric context color histogram features	17,032
Elements	[8]	low	color: mean saturation, brightness and hue, emotional coordinates, colorfulness, color names, Itten contrast, Wang's semantic descriptions of colors, area statistics; texture: Tamura, Wavelet and gray-level co-occurrence matrix	97
MPEG-7	[65]	low	color: layout, structure, scalable color, dominant color; texture: edge histogram, texture browsing	≈200
Shape	[92]	low	line segments, continuous lines, angles, curves	219
IttenColor	[93]	low	color co-occurrence features and patch-based color-combination features	16,485
Attributes	[7]	mid	scene attributes	102
Sentributes	[94]	mid	scene attributes, eigenfaces	109
Constructs	[95]	mid	roundness, angularity, complexity	3
Composition	[8]	mid	level of detail, low depth of field, dynamics, rule of thirds	45
Aesthetics	[96]	mid	figure-ground relationship, color pattern, shape, composition	13
Principles	[9]	mid	principles-of-art: balance, contrast, harmony, variety, gradation, movement	165
SIFT	[97]	mid	bag-of-visual-words on SIFT, latent topics	330
FS	[8]	high	number of faces and skin pixels, size of the biggest face, amount of skin w.r.t. the size of faces	4
ANP	[10]	high	semantic concepts based on adjective noun pairs	1,200
Expressions	[98]	high	automatically assessed facial expressions (anger, contempt, disgust, fear, happiness, sadness, surprise, neutral)	8
HLCs	[99]	high	object information and scene information	1,205

Affective gap examples

➢ Different perspectives of deep features:
- Combining local and global features, such as WSCNet.
- Extracting multi-level features, such as MldrNet, CNN-RNN.
- Considering emotional polarity, such as RCA, APSE.
- For domain adaption, such as EmotionGAN, MSGAN.

Future

- Based on the **understanding** of images, considering the context of the image may be helpful.
- Based on the **cost** of data collection, learning from noisy data or few labels may be more useful.
- Based on the **subjective** of emotion, creating a large-scale dataset with high-quality annotation, may be significantly advance the development of AICA.
- Based on the **reality** of research, some novel and real-world AICA-based applications may bring more chances.



The importance of context

The example of subjective

Minsky (a Turing Award winner in 1970) claimed that "The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without emotions." An emotional intelligence era with more AICA-based real-world applications is coming.

Datasets

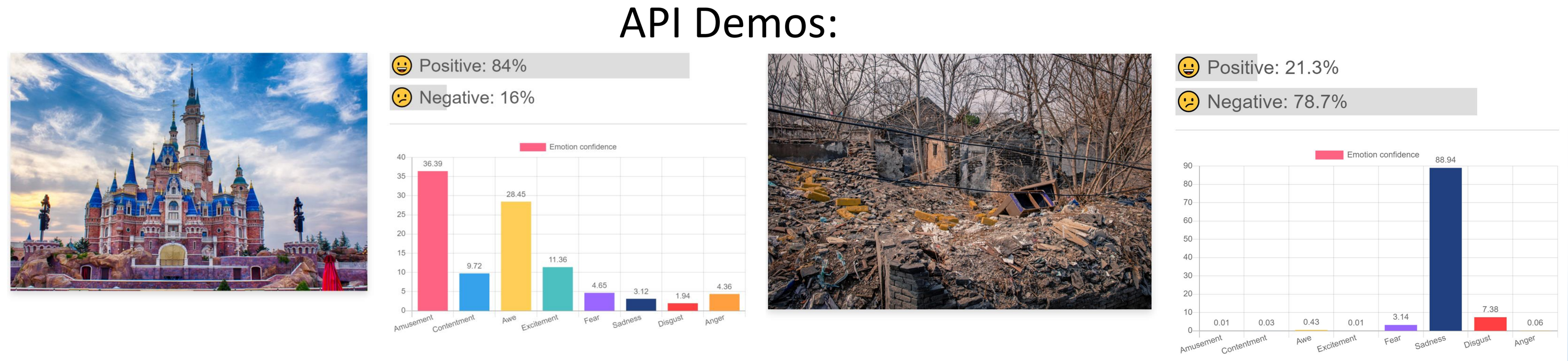
➢ Different Tasks of datasets:
- **Dominant emotion recognition**, such as FI, IAPSa, Artphoto, Abstract, etc.
- **Emotion distribution learning**, such as Emotion6, Flickr_LDL and Twitter_LDL.
- **Personalized emotion prediction**, such as IESN.
- **Learning from noisy data**, such as WEBEmo, StockEmotion.

Dataset	Ref	# Images	Type	# Annotators	Emotion model	Label detail
IAPS	[28]	1,182	natural	≈100 (half f)	VAD	empirically derived mean and standard deviation
IAPSa	[29]	390	natural	20 (10f,10m)	Mikels	at least one emotion category for each image
Abstract	[8]	280	abstract	≈230	Mikels	the detailed votes of all emotions for each image
ArtPhoto	[8]	806	artistic	–	Mikels	one DEC for each image
GAPED	[74]	730	natural	60	Sentiment, VA	one DEC and average VA values for each image
MART	[75]	500	abstract	25 (11f,14m)	Sentiment	one DEC for each image
devArt	[75]	500	abstract	60 (27f,33m)	Sentiment	one DEC for each image
Twitter I	[76]	1,269	social	5 per image	Sentiment	one sentiment category for each image
Twitter II	[10]	603	social	3 per image	Sentiment	one sentiment category for each image
VSO	[10]	≈500,000	social	–	Plutchik	one emotion category for each image
MVSO	[77]	7.36M	social	–	Plutchik	one emotion category for each image
Flickr I	[78]	354,192	social	6,735	Ekman	one emotion category for each image
Flickr II	[79]	60,745	social	3 per image	Sentiment	one sentiment category for each image
Instagram	[79]	42,856	social	3 per image	Sentiment	one sentiment category for each image
Emotion6	[14]	1,980	social	432	Ekman+neutral, VA	the discrete probability distribution
FI	[4]	23,308	social	225	Mikels	one DEC for each image
IESN	[15]	1,012,901	social	118,035	Mikels, VAD	the emotion of involved users for each image
T4SA	[80]	1,473,394	social	-	Sentiment+neutral	one sentiment category for each image
B-T4SA	[80]	470,586	social	-	Sentiment+neutral	one sentiment category for each image
Comics	[81]	11,821	comic	10 (5f,5m)	Mikels	one DEC for each image
Event	[82]	8,748	social	3 each image	Sentiment+neutral	one sentiment category for each image
EMOTIC	[83]	18,316	social	3 each image	Ekman, VAD	one DEC and VAD values for each image
EMOd	[84]	1,019	natural	3	Sentiment+neutral	object contour, object name, sentiment category
WEBEmo	[22]	268,000	social	-	Parrott	one DEC for each image
LUCFER	[85]	3.6M	social	-	Plutchik, VAD, context	one DEC, average VAD values, and context for each image
FlickrLDL	[16]	10,700	social	11	Mikels	the discrete probability distribution
TwitterLDL	[16]	10,045	social	8	Mikels	the discrete probability distribution

More Details

Download paper: Our website:

API Demos:



Experiments

Dataset	PAEF [9]				Sun attribute [7]				SentiBank [10]			
	kNN	NB	SVM	Avg	kNN	NB	SVM	Avg	kNN	NB	SVM	Avg
Emotion6	0.246	0.288	0.359	0.298	0.268	0.323	0.306	0.299	0.283	0.290	0.342	0.305
FI2	0.687	0.733	0.730	0.717	0.698	0.697	0.739	0.711	0.603	0.815	0.815	0.744
FI8	0.286	0.299	0.343	0.309	0.300	0.271	0.372	0.314	0.445	0.288	0.506	0.413
Flickr	0.627	0.640	0.674	0.647	0.634	0.639	0.683	0.652	0.581	0.608	0.694	0.628
Instagram	0.556	0.589	0.638	0.594	0.561	0.586	0.631	0.593	0.584	0.576	0.662	0.607
Twitter I	0.593	0.633	0.675	0.634	0.565	0.615	0.643	0.608	0.526	0.564	0.602	0.564
Twitter II	0.659	0.777	0.777	0.738	0.672	0.606	0.777	0.685	0.632	0.661	0.777	0.690

Note that Sun attribute represent low-level feature here.

➢ Comparison of different types of hand-crafted-features:
- On large-scale datasets, high level features have best performance.

➢ Comparison between hand-crafted features and deep features:
- Deep features are much better than hand-crafted features.

Dataset	WSCNet [12]			PDANet [111]		
	L	G	L+G	L	G	L+G
FI2	0.894 (3)	0.894 (3)	0.896 (1)	0.807 (3)	0.876 (2)	0.878 (1)
FI8	0.671 (3)	0.675 (2)	0.679 (1)	0.606 (3)	0.696 (1)	0.694 (2)
Flickr_LDL	0.697 (3)	0.707 (2)	0.709 (1)	0.592 (3)	0.703 (1)	0.703 (1)
Twitter_LDL	0.764 (3)	0.773 (1)	0.766 (2)	0.725 (3)	0.762 (2)	0.763 (1)
Comics	0.531 (3)	0.532 (2)	0.542 (1)	0.263 (3)	0.595 (1)	0.588 (2)
GAPED	0.899 (2)	0.889 (3)	0.919 (1)	0.697 (3)	0.939 (2)	0.950 (1)
Event	0.938 (2)	0.937 (3)	0.948 (1)	0.791 (3)	0.937 (2)	0.946 (1)
Flickr	0.800 (3)	0.801 (2)	0.807 (1)	0.757 (3)	0.808 (2)	0.819 (1)
Instagram	0.804 (3)	0.816 (1)	0.804 (3)	0.672 (3)	0.811 (1)	0.807 (2)
Twitter I	0.819 (3)	0.827 (1)	0.827 (1)	0.606 (3)	0.839 (2)	0.858 (1)
Twitter II	0.824 (1)	0.824 (1)	0.815 (3)	0.815 (1)	0.815 (1)	0.815 (1)
Average rank	2.636 (3)	1.909 (2)	1.455 (1)	2.818 (3)	1.545 (2)	1.273 (1)

➢ Comparison of different type of deep features:
- Global features better than local features in general.
- Combining local and global features have best performance.

➢ Noise:
- Least: Event
- Most: Abstract

➢ Observations:
- Different types
- Not mutual
- Class distribution
- FI dataset

Dataset	ρ_{-1}	ρ_{+1}	ρ_m
Event	0.148	0.056	0.074
Twitter_LDL	0.404	0.068	0.097
WEBEmo	0.098	0.115	0.107
FI	0.169	0.082	0.108
Flickr_LDL	0.278	0.098	0.129
Comics	0.171	0.189	0.180
Flickr	0.184	0.176	0.180
GAPED	0.170	0.227	0.187
Instagram	0.195	0.179	0.187
Twitter I	0.290	0.156	0.209
Artphoto	0.245	0.271	0.257
Twitter II	0.492	0.198	0.264
IAPSa	0.315	0.257	0.283
Abstract	0.387	0.257	0.307

(a) Label noise

	Event	Twitter_LDL	WEBEmo	FI	Flickr_LDL	Comics	Flickr	GAPED	Instagram	Twitter I	Artphoto	Twitter II	IAPSa	Abstract
Event	0.931	0.841	0.452	0.694	0.753	0.49	0.587	0.515	0.556	0.795	0.503	0.849	0.532	0.674
Twitter_LDL	0.793	0.918	0.444	0.732	0.842	0.542	0.519	0.404	0.523	0.634	0.596	0.832	0.675	0.609
WEBEmo	0.75	0.563	0.804	0.7	0.573	0.601	0.677	0.747	0.669	0.713	0.683	0.681	0.74	0.609
FI	0.752	0.673	0.602	0.875	0.729	0.609	0.722	0.636	0.69	0.65	0.764	0.647	0.818	0.739
Flickr_LDL	0.787	0.905	0.448	0.754	0.882	0.582	0.555	0.566	0.556	0.61	0.621	0.824	0.701	0.587
Comics	0.707	0.814	0.465	0.706	0.721	0.813	0.562	0.545	0.601	0.614	0.627	0.739	0.766	0.522
Flickr	0.842	0.664	0.501	0.769	0.657	0.595	0.804	0.677	0.731	0.78	0.795	0.647	0.727	0.652
GAPED	0.624	0.589	0.472	0.605	0.616	0.584	0.585	0.919	0.585	0.575	0.571	0.63	0.662	0.609
Instagram	0.853	0.747	0.495	0.777	0.708	0.592	0.786	0.727	0.784	0.803	0.745	0.748	0.779	0.696
Twitter I	0.882	0.867	0.455	0.735	0.803	0.494	0.616	0.505	0.582	0.823	0.534	0.815	0.662	0.652
Artphoto	0.717	0.584	0.494	0.677	0.469	0.575	0.702	0.747	0.628	0.571	0.807	0.613	0.74	0.739
Twitter II	0.826	0.908	0.437	0.7	0.816	0.47	0.505	0.333	0.498	0.654	0.491	0.824	0.532	0.63
IAPSa	0.762	0.83	0.464	0.734	0.759	0.519	0.549	0.727	0.551	0.543	0.584	0.739	0.818	0.674
Abstract	0.738	0.646	0.528	0.644	0.591	0.476	0.586	0.646	0.573	0.622	0.665	0.639	0.571	0.739

(b) Dataset bias