



Towards Hardware-Aware Sentiment Analysis

Edoardo Ragusa
DITEN, University of Genova,
Italy

WISDOM'20
(KDD 2020, August 24th, San Diego)



Outline

- Introduction
 - Embedded systems for sentiment analysis on the edge
- Image polarity detection on the edge
- Cognitive models and computational resources



Introduction

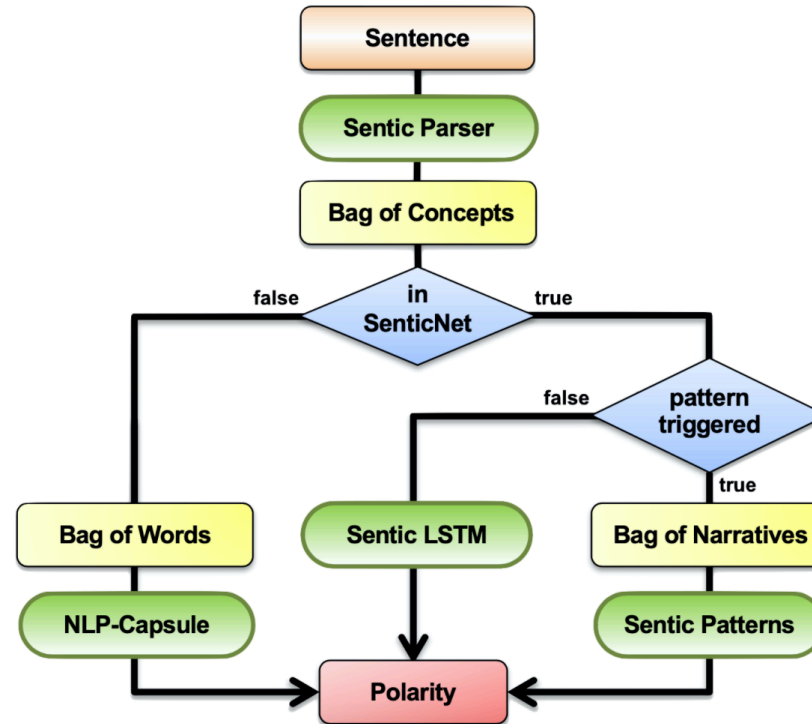


Sentiment analysis

- Sentiment analysis infers users' emotions automatically
 - Traditional artificial intelligence
 - Small computing power
 - Handcrafted feature sets
 - Deep learning
 - High computing power
 - Automatic feature learning



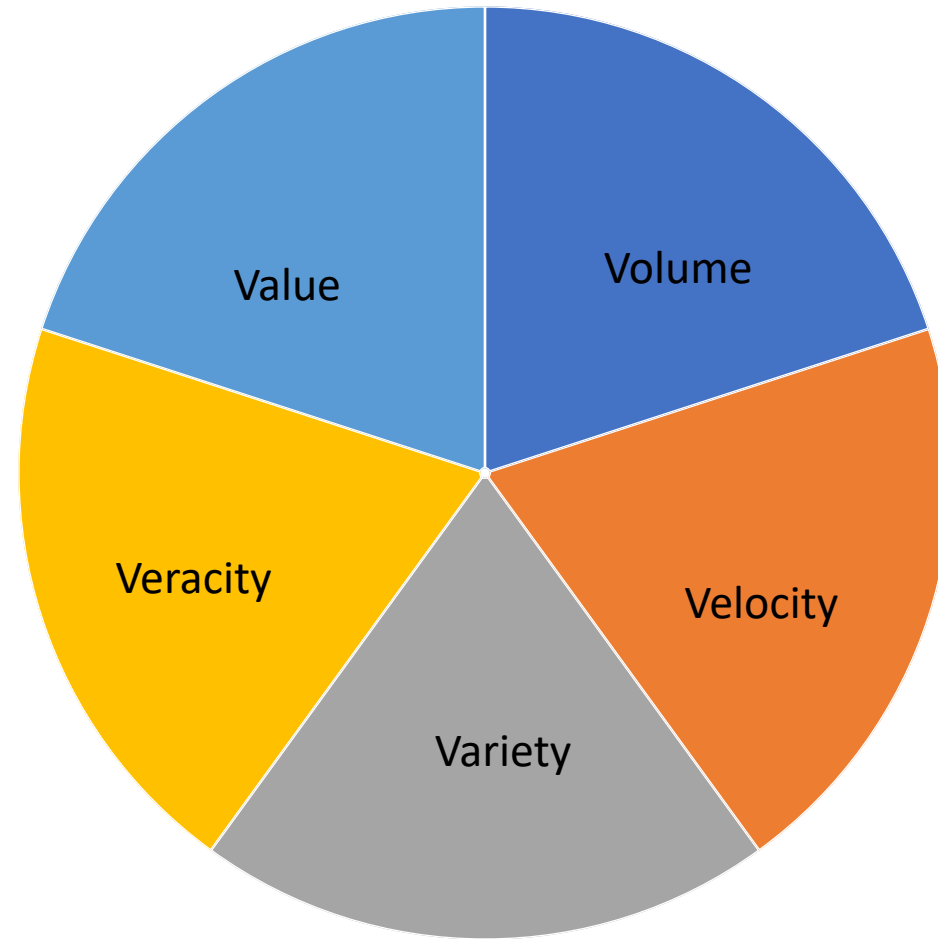
Sentiment analysis: text



Cambria, E., & Hussain, A. (2012). Sentic computing. *marketing*, 59(2), 557-577.

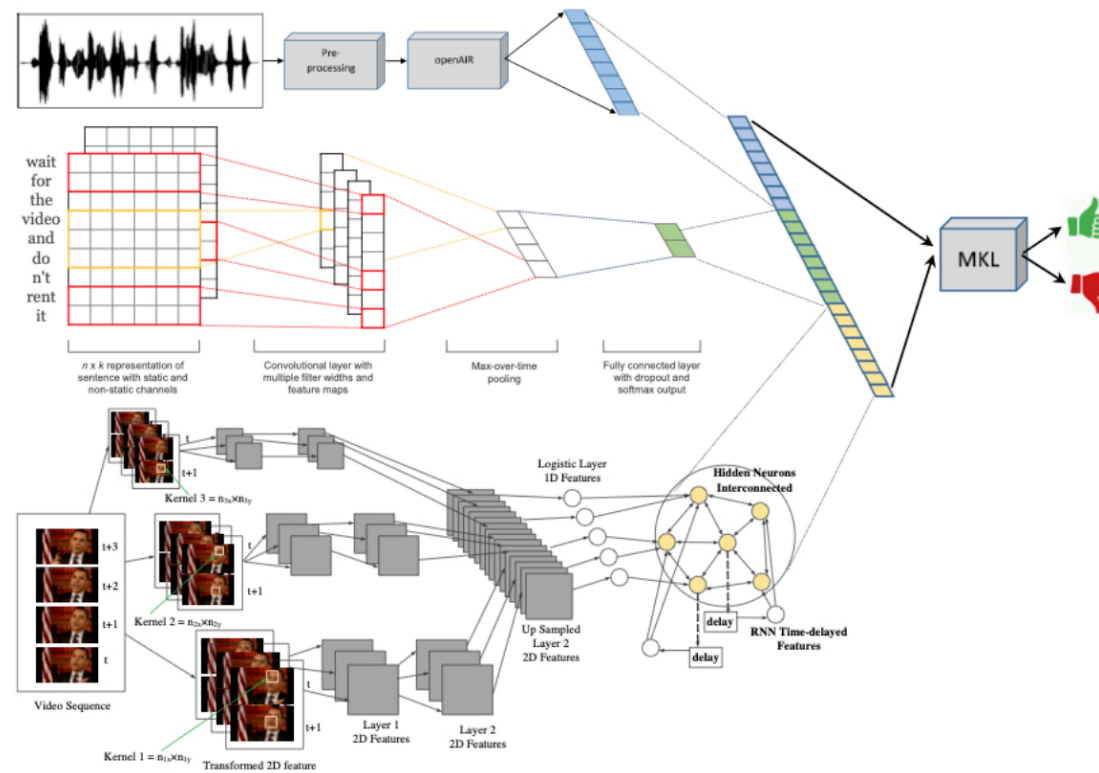


Big data era





Sentiment Analysis: multimodal



Poria, S., Chaturvedi, I., Cambria, E., & Hussain, A. (2016, December). Convolutional MKL based multimodal emotion recognition and sentiment analysis. In *2016 IEEE 16th international conference on data mining (ICDM)* (pp. 439-448). IEEE.



Sentiment Analysis: hardware

- Smart
 - Phones
 - Watches
 - Speakers
 - Cameras
 - TVs
 - Glasses
 - Cars
 - Refrigerators



Sentiment Analysis: hardware

- Platforms for deep learning on embedded devices:
 - ***Nvidia Jetson***
 - Movidius
 - Google Colab
 - STM32
 - NPUs
 - VPUs





Sentiment Analysis: hardware

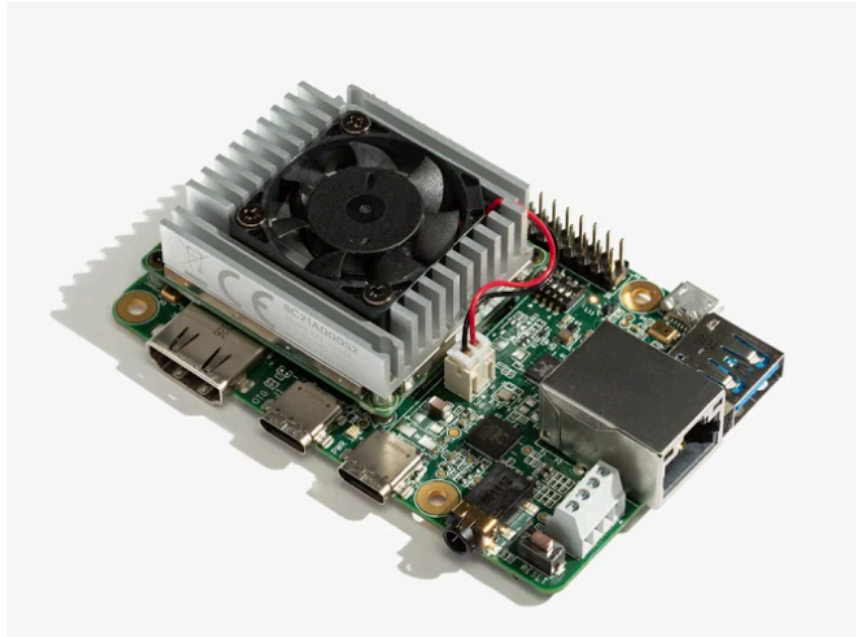
- Platforms for deep learning on embedded devices:
 - Nvidia Jetson
 - **Movidius**
 - Google Colab
 - STM32
 - NPUs
 - VPUs





Sentiment Analysis: hardware

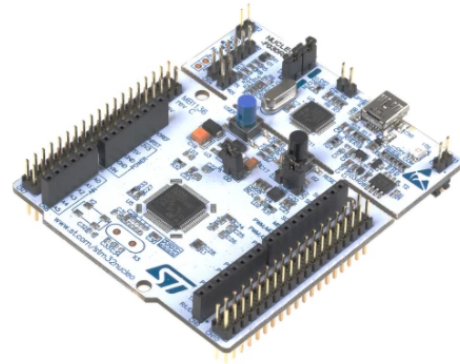
- Platforms for deep learning on embedded devices:
 - Nvidia Jetson
 - Movidius
 - **Google Colab**
 - STM32
 - NPUs
 - VPUs





Sentiment Analysis: hardware

- Platforms for deep learning on embedded devices:
 - Nvidia Jetson
 - Movidius
 - Google Colab
 - **STM32**
 - NPUs
 - VPUs





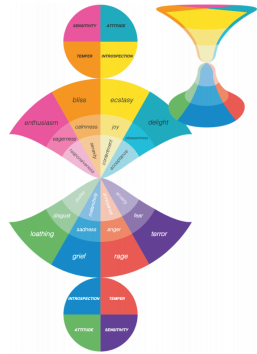
Sentiment Analysis: hardware

- Platforms for deep learning on embedded devices:
 - Nvidia Jetson
 - Movidius
 - Google Colab
 - STM32
 - ***NPU***s
 - ***VPU***s

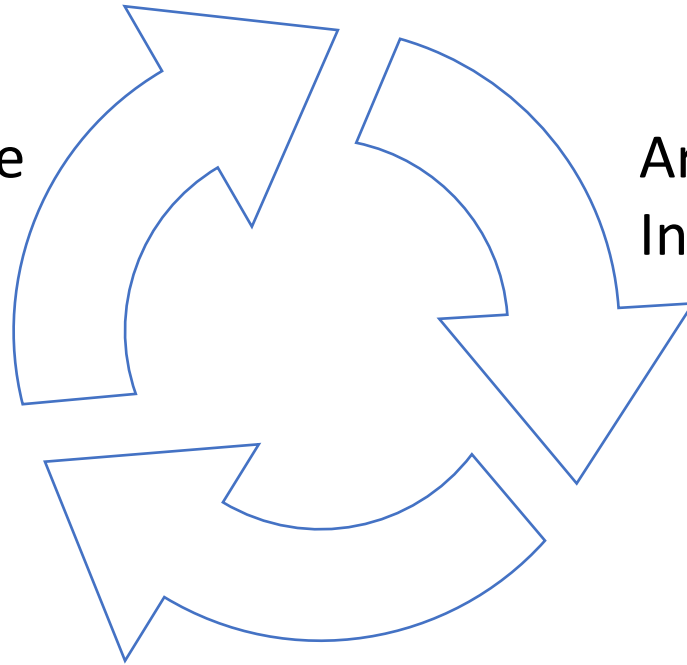




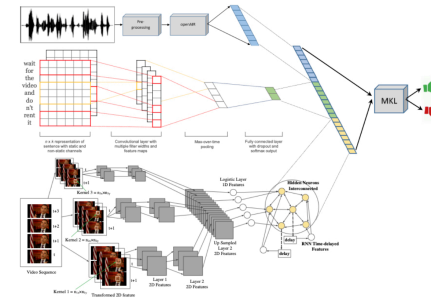
Hardware-algorithm loop



Cognitive models



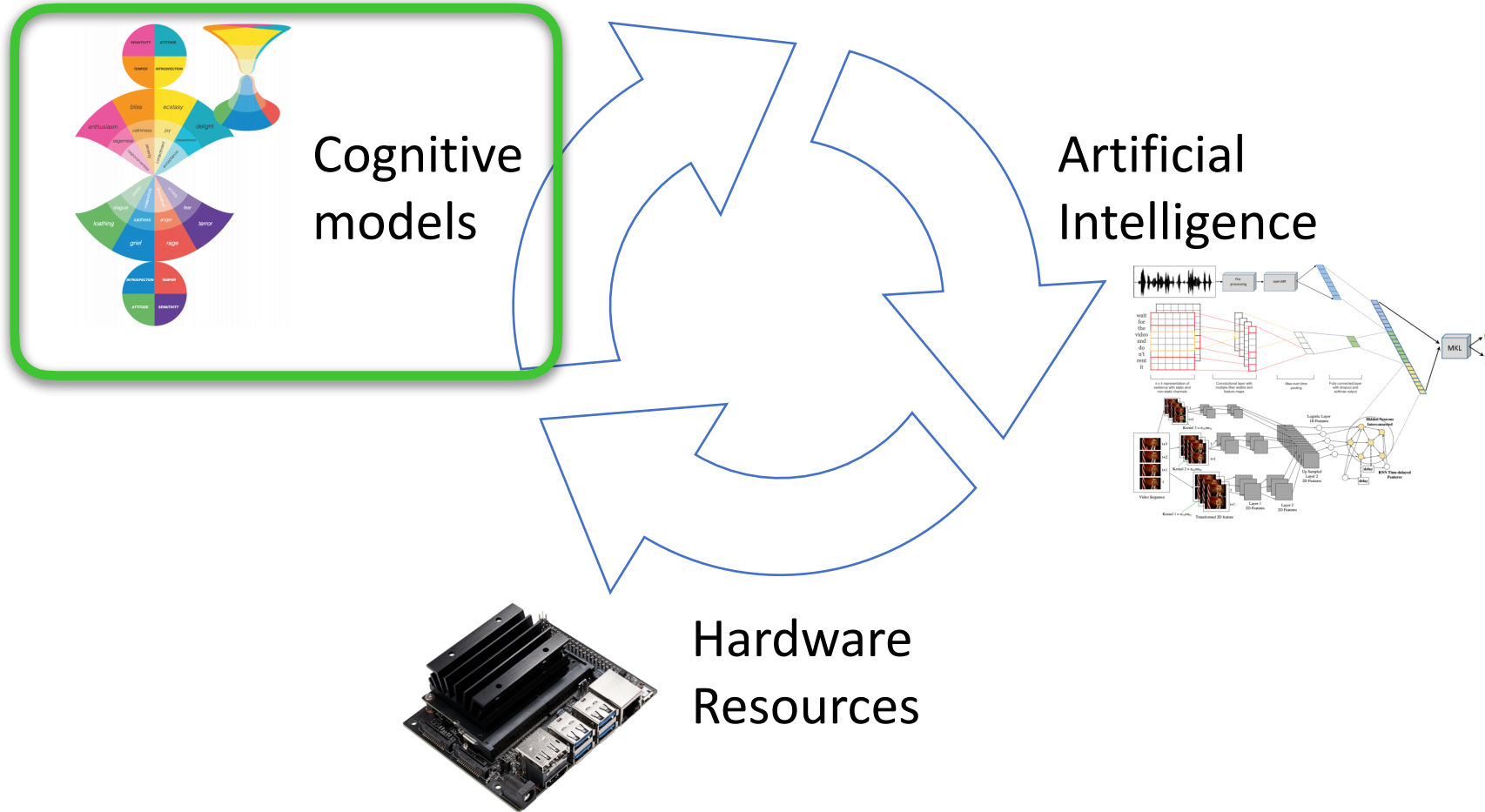
Artificial Intelligence



Hardware Resources

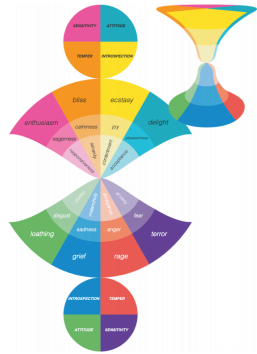


Hardware-algorithm loop

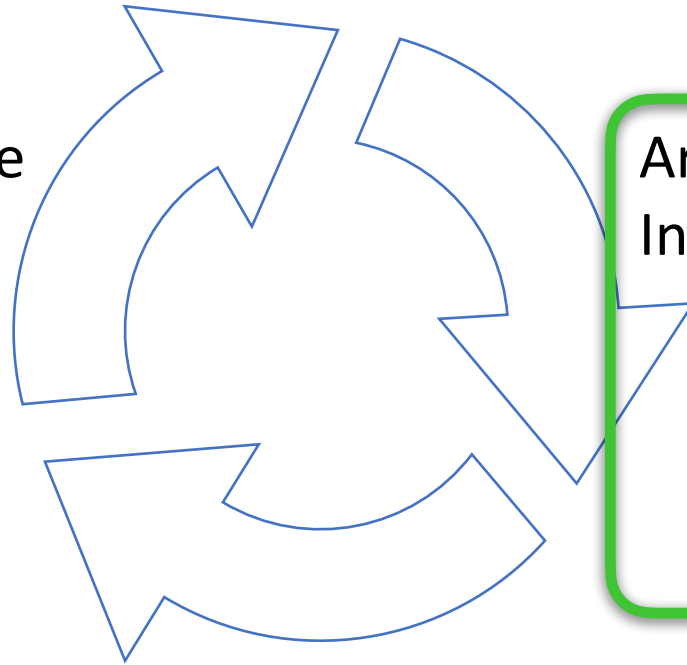




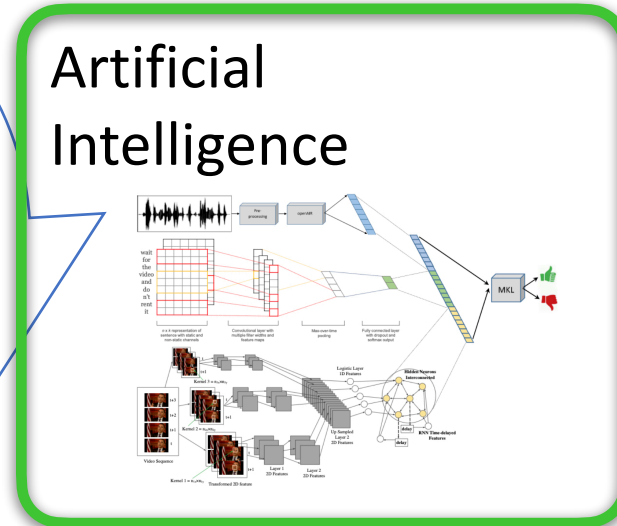
Hardware-algorithm loop



Cognitive models

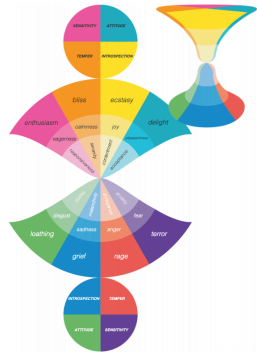


Hardware Resources

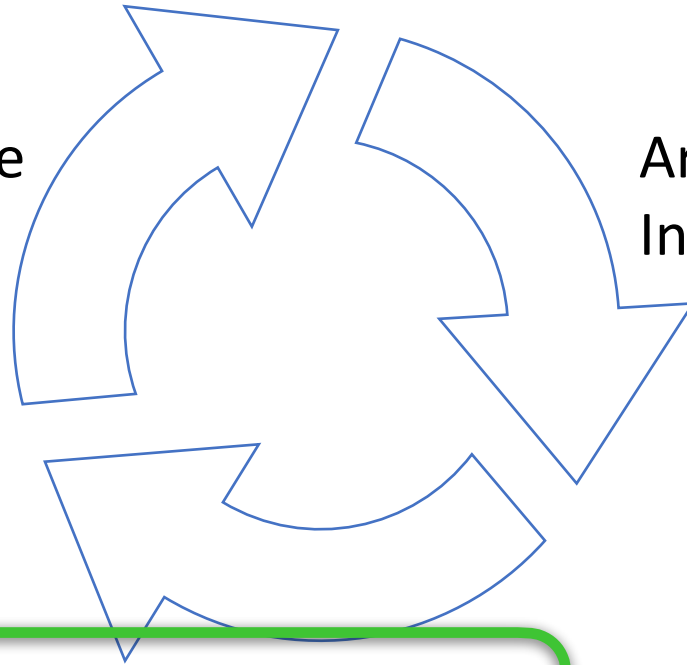




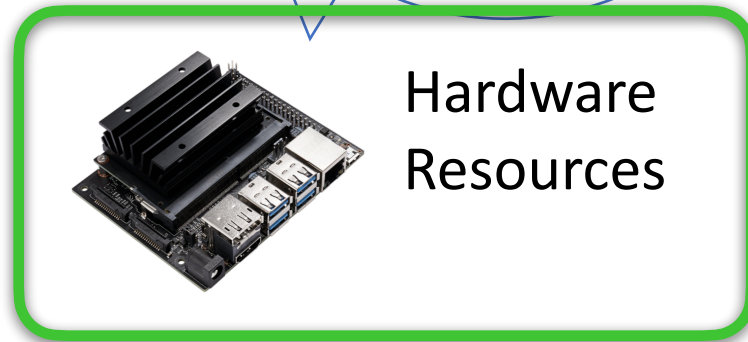
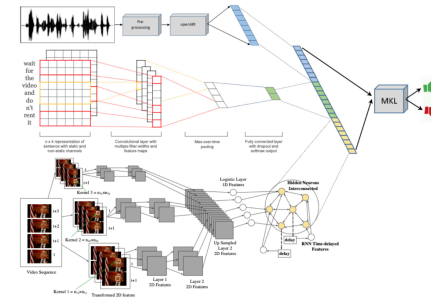
Hardware-algorithm loop



Cognitive models



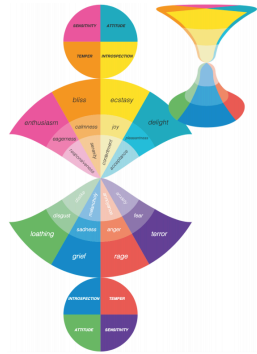
Artificial Intelligence



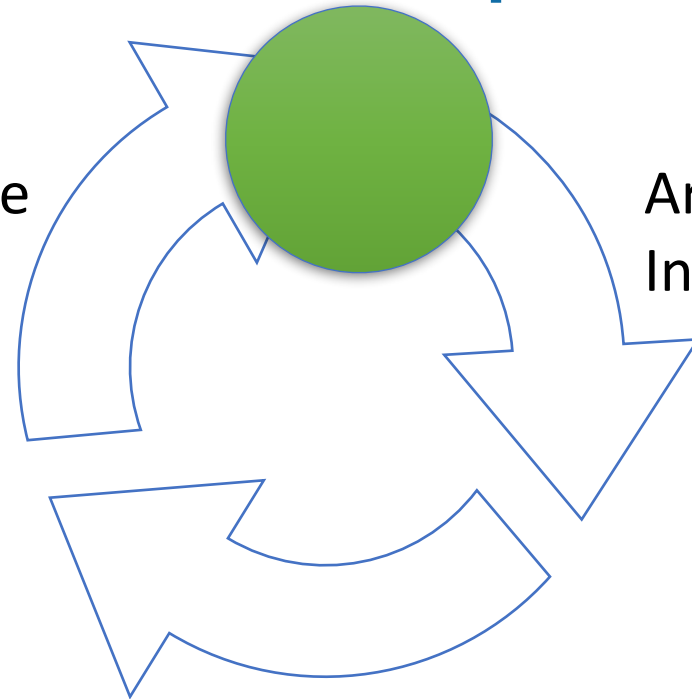
Hardware Resources



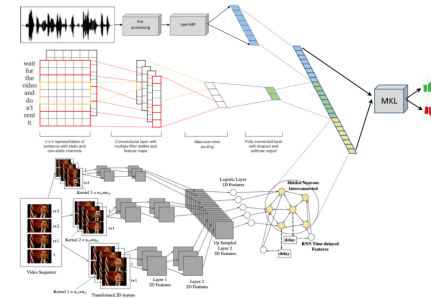
Hardware-algorithm loop



Cognitive models



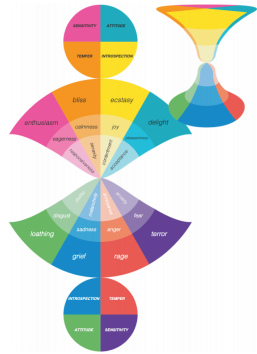
Artificial Intelligence



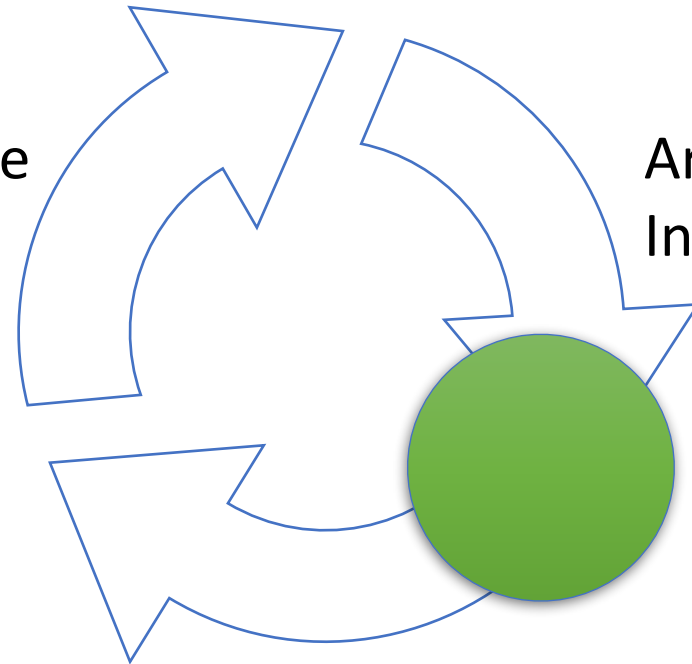
Hardware Resources



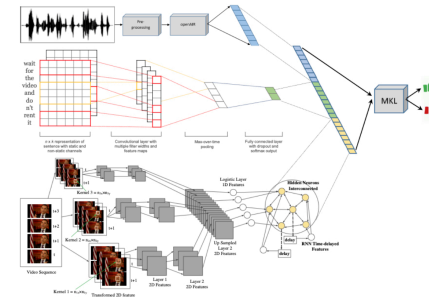
Hardware-algorithm loop



Cognitive models



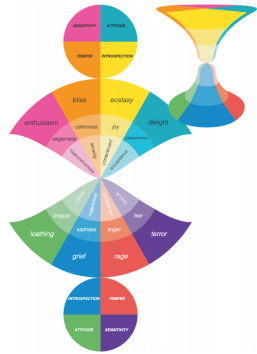
Artificial Intelligence



Hardware Resources

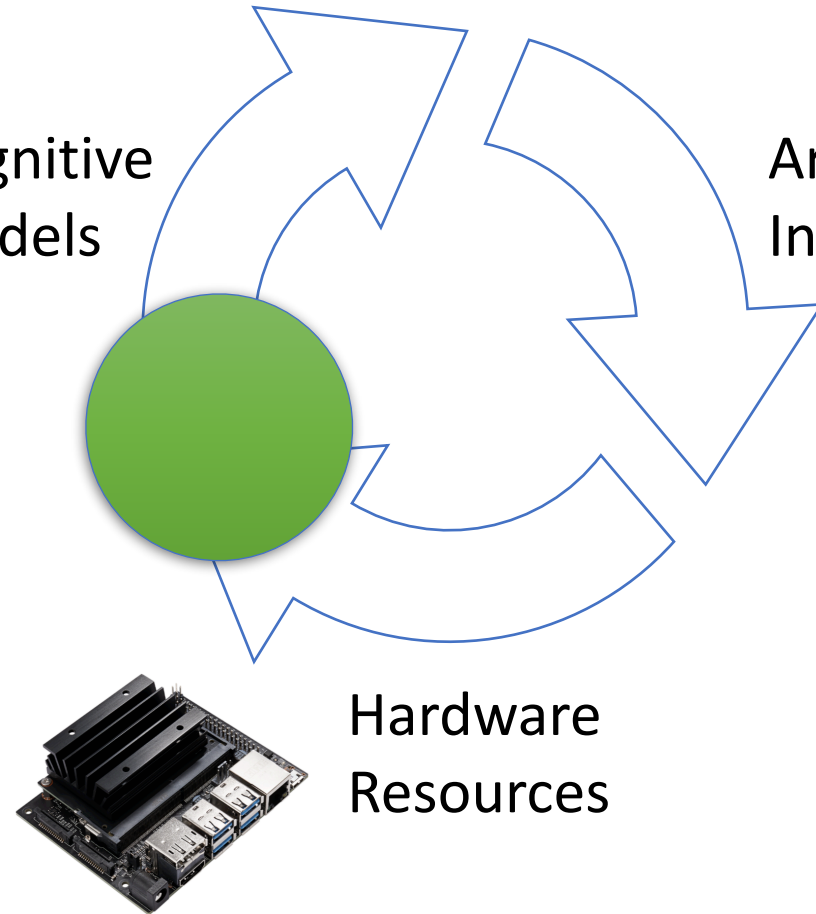


Hardware-algorithm loop

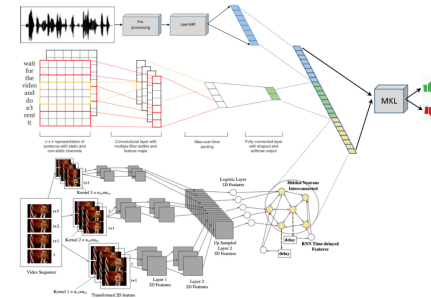


Cognitive models

Artificial Intelligence

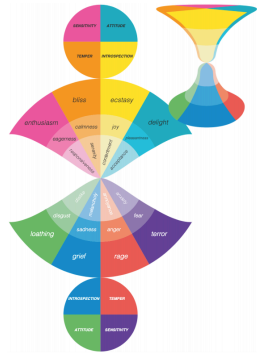


Hardware Resources

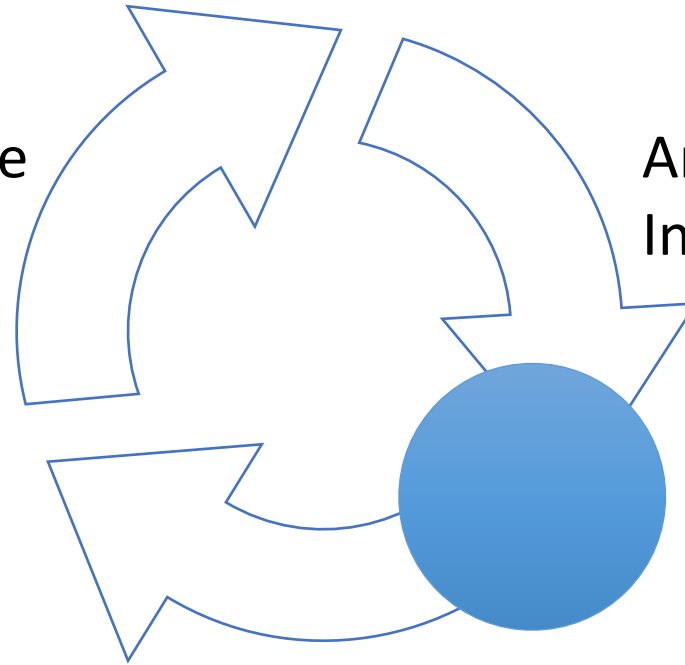




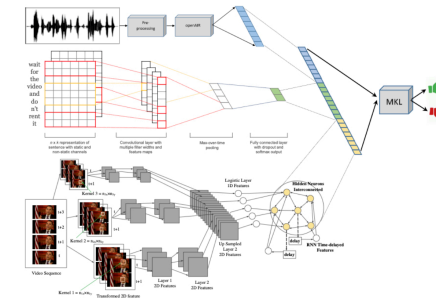
Hardware-algorithm loop



Cognitive models



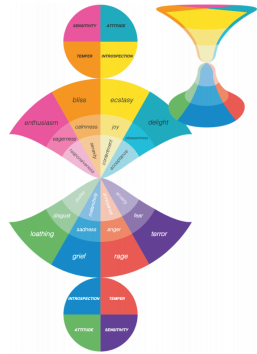
Artificial Intelligence



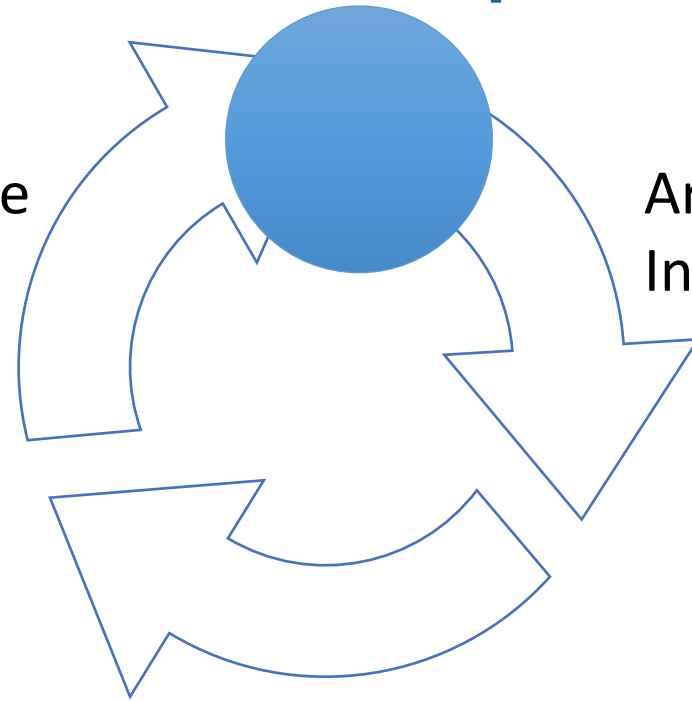
Hardware Resources



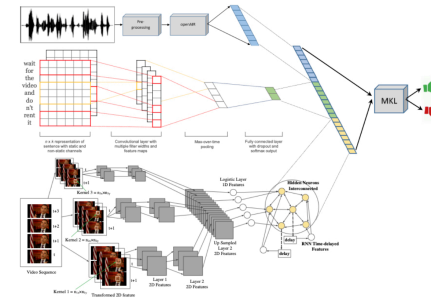
Hardware-algorithm loop



Cognitive models



Artificial Intelligence



Hardware Resources



Image polarity detection on the edge



Image polarity detection

- ***Polarity***, also known as orientation is the emotion expressed in a content. Usually, it is expressed as positive, negative or neutral.



Image polarity detection

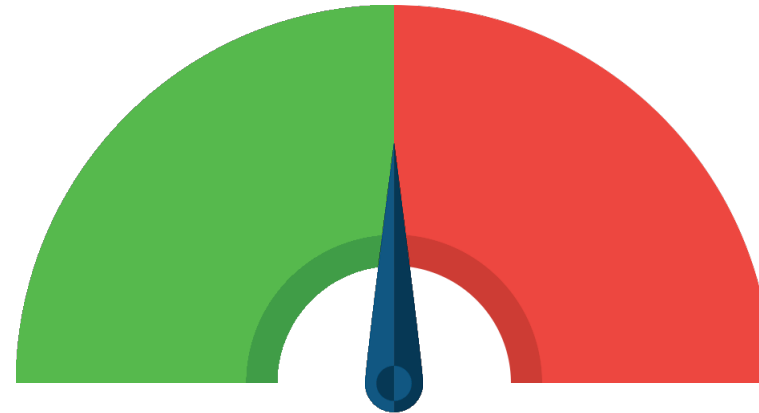




Image polarity detection



POSITIVE





Image polarity detection

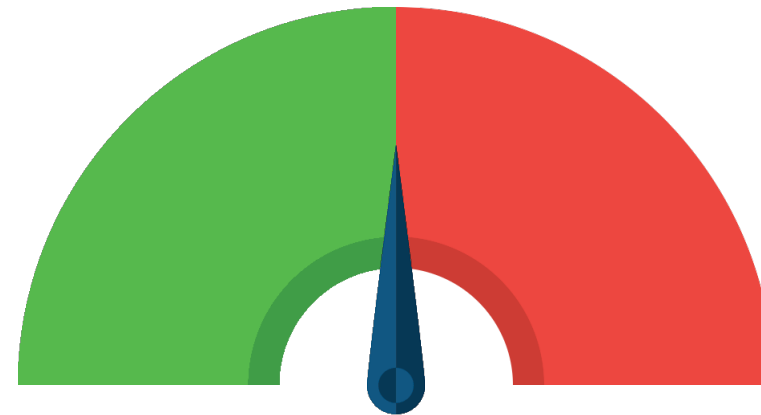
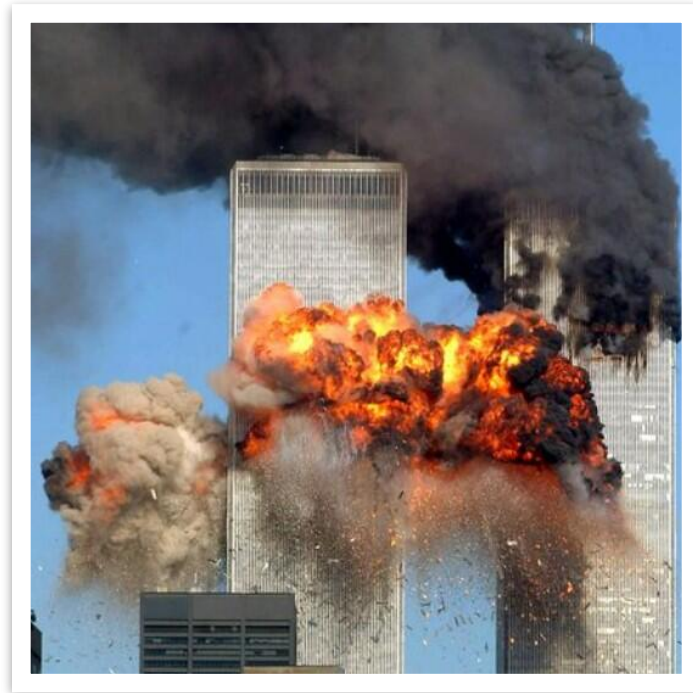
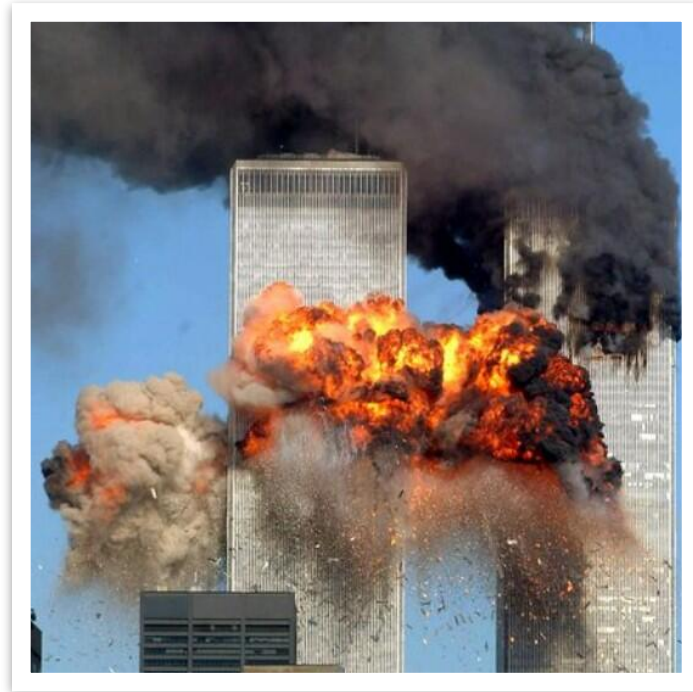




Image polarity detection



NEGATIVE

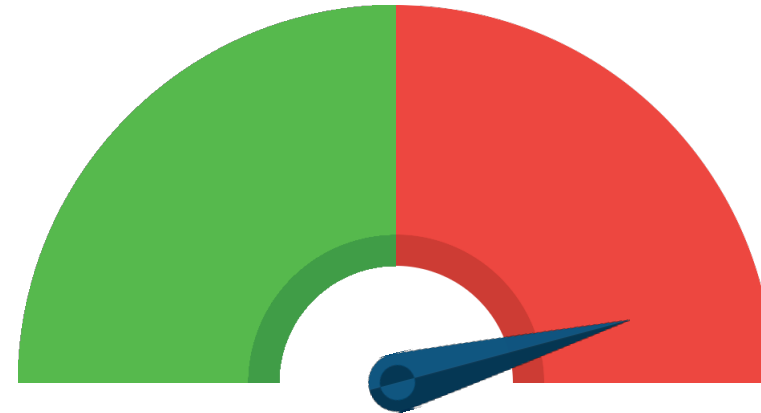
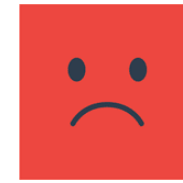




Image polarity detection

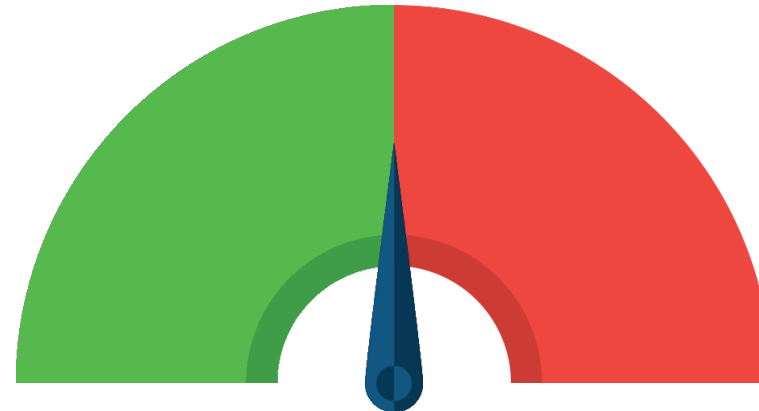
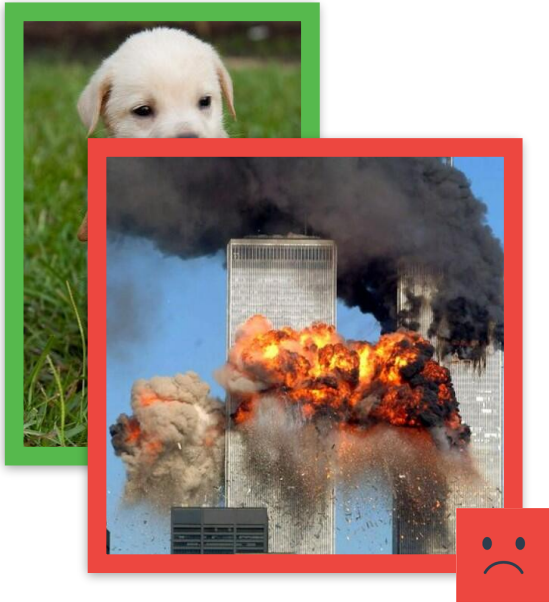




Image polarity detection

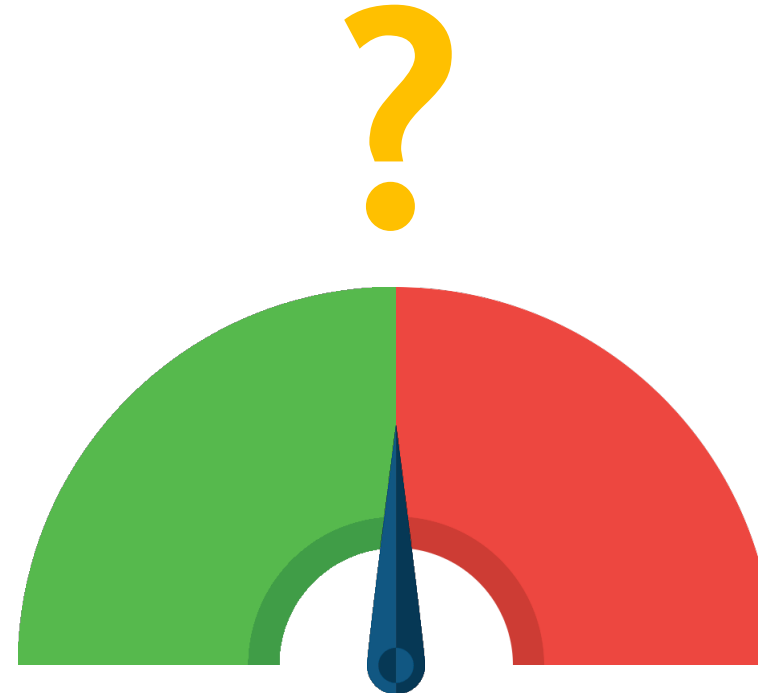
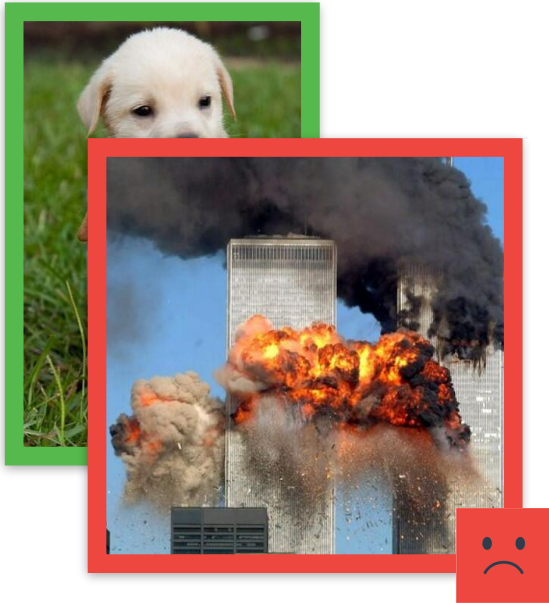
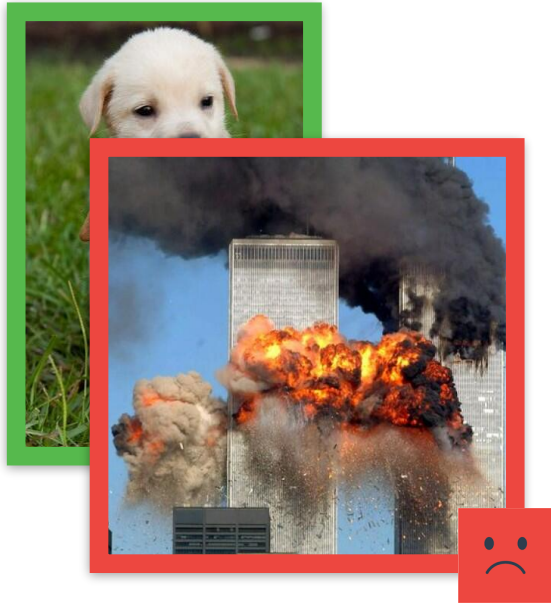




Image polarity detection



Applications

- Rehabilitation
- Health care systems
- Driving assistance
- Human robot interaction
- Security
- Aiding devices





Why sentiments analysis on the edge

- Cloud computing
 - Strong internet connections
 - High performance hardware on server-side
 - Privacy concerns
 - sensitive information
 - low information about data storage and management
- Smart devices
 - Real-time
 - Low power
 - Safer from the privacy point of view



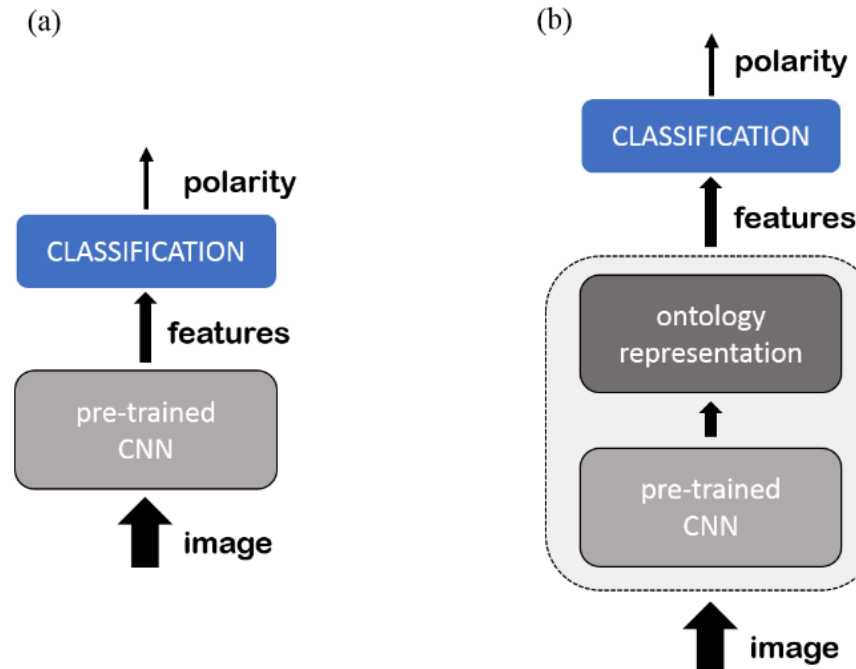
CNNs

- State-of-the-art for all image processing tasks:
 - Image Classification
 - Object Detection
- Drawbacks:
 - Training phase
 - Hardware deployment
 - Power-hungry
 - Memory demanding



Baseline approach

- Fine tune object detection architectures



Ragusa, E., Cambria, E., Zunino, R., & Gastaldo, P. (2019). A survey on deep learning in image polarity detection: Balancing generalization performances and computational costs. *Electronics*, 8(7), 783.



Object classification architecture

Architecture	Weight Layers	Acc (%)	Operations (Gflops)	Parameters ($\cdot 10^6$)
AlexNet [26]	8	54	0.7	61
Vgg_16 [27]	16	71	15.5	138
Vgg_19 [27]	19	71	19.6	144
GoogLeNet [28]	58	68	1.6	7
Inc_v3 [29]	46	78	6	24
Res_50 [31]	50	76	3.9	26
Res_101 [31]	101	77	7.6	45
Res_152 [31]	152	79	11.3	60
DenseNet [32]	201	77	4.0	20

Ragusa, E., Cambria, E., Zunino, R., & Gastaldo, P. (2019). A survey on deep learning in image polarity detection: Balancing generalization performances and computational costs. *Electronics*, 8(7), 783.



Object classification architecture

Architecture	Weight Layers	Acc (%)	Operations (Gflops)	Parameters ($\cdot 10^6$)
AlexNet [26]	8	54	0.7	61
Vgg_16 [27]	16	71	15.5	138
Vgg_19 [27]	19	71	19.6	144
GoogLeNet [28]	58	68	1.6	7
Inc_v3 [29]	46	78	6	24
Res_50 [31]	50	76	3.9	26
Res_101 [31]	101	77	7.6	45
Res_152 [31]	152	79	11.3	60
DenseNet [32]	201	77	4.0	20

Ragusa, E., Cambria, E., Zunino, R., & Gastaldo, P. (2019). A survey on deep learning in image polarity detection: Balancing generalization performances and computational costs. *Electronics*, 8(7), 783.



Object classification architecture

Architecture	Weight Layers	Acc (%)	Operations (Gflops)	Parameters ($\cdot 10^6$)
AlexNet [26]	8	54	0.7	61
Vgg_16 [27]	16	71	15.5	138
Vgg_19 [27]	19	71	19.6	144
GoogLeNet [28]	58	68	1.6	7
Inc_v3 [29]	46	78	6	24
Res_50 [31]	50	76	3.9	26
Res_101 [31]	101	77	7.6	45
Res_152 [31]	152	79	11.3	60
DenseNet [32]	201	77	4.0	20

Ragusa, E., Cambria, E., Zunino, R., & Gastaldo, P. (2019). A survey on deep learning in image polarity detection: Balancing generalization performances and computational costs. *Electronics*, 8(7), 783.



Object classification architecture

Name	Size	Positive Samples	Negative Samples
Tw [12]	882	581	301
Mv_a [74]	702	351	351
Mv_b [74]	702	351	351
ANP40 [22]	17114	11857	5257

Ragusa, E., Cambria, E., Zunino, R., & Gastaldo, P. (2019). A survey on deep learning in image polarity detection: Balancing generalization performances and computational costs. *Electronics*, 8(7), 783.



Object classification polarity detection

Architecture	Tw(882)	Mv_a(702)	Mv_b(702)	ANP40(17K)
AlexNet	82.5 (0.6)	66.2 (0.5)	65.4 (0.7)	76.3 (0.3)
Vgg_16	86.6 (0.4)	68.9 (1.0)	70.7 (1.0)	79.0 (0.3)
Vgg_19	86.4 (0.5)	69.2 (0.7)	71.0 (0.8)	78.7 (0.3)
GoogLeNet	84.2 (0.5)	66.0 (0.6)	68.2 (0.5)	79.2 (0.3)
Inc_v3	86.5 (0.7)	68.1 (0.7)	70.5 (0.9)	79.9 (0.2)
Res_50	85.8 (0.6)	68.9 (0.9)	68.8 (1.5)	79.2 (0.2)
Res_101	88.2 (0.4)	70.8 (0.6)	69.2 (0.8)	79.7 (0.2)
DenseNet	89.4 (0.5)	71.3 (0.8)	70.6 (0.6)	79.3 (0.3)

Ragusa, E., Cambria, E., Zunino, R., & Gastaldo, P. (2019). A survey on deep learning in image polarity detection: Balancing generalization performances and computational costs. *Electronics*, 8(7), 783.



Object classification polarity detection

Architecture	Tw(882)	Mv_a(702)	Mv_b(702)	ANP40(17K)
AlexNet	82.5 (0.6)	66.2 (0.5)	65.4 (0.7)	76.3 (0.3)
Vgg_16	86.6 (0.4)	68.9 (1.0)	70.7 (1.0)	79.0 (0.3)
Vgg_19	86.4 (0.5)	69.2 (0.7)	71.0 (0.8)	78.7 (0.3)
GoogLeNet	84.2 (0.5)	66.0 (0.6)	68.2 (0.5)	79.2 (0.3)
Inc_v3	86.5 (0.7)	68.1 (0.7)	70.5 (0.9)	79.9 (0.2)
Res_50	85.8 (0.6)	68.9 (0.9)	68.8 (1.5)	79.2 (0.2)
Res_101	88.2 (0.4)	70.8 (0.6)	69.2 (0.8)	79.7 (0.2)
DenseNet	89.4 (0.5)	71.3 (0.8)	70.6 (0.6)	79.3 (0.3)

Ragusa, E., Cambria, E., Zunino, R., & Gastaldo, P. (2019). A survey on deep learning in image polarity detection: Balancing generalization performances and computational costs. *Electronics*, 8(7), 783.



Object classification polarity detection

Architecture	Tw(882)	Mv_a(702)	Mv_b(702)	ANP40(17K)
AlexNet	82.5 (0.6)	66.2 (0.5)	65.4 (0.7)	76.3 (0.3)
Vgg_16	86.6 (0.4)	68.9 (1.0)	70.7 (1.0)	79.0 (0.3)
Vgg_19	86.4 (0.5)	69.2 (0.7)	71.0 (0.8)	78.7 (0.3)
GoogLeNet	84.2 (0.5)	66.0 (0.6)	68.2 (0.5)	79.2 (0.3)
Inc_v3	86.5 (0.7)	68.1 (0.7)	70.5 (0.9)	79.9 (0.2)
Res_50	85.8 (0.6)	68.9 (0.9)	68.8 (1.5)	79.2 (0.2)
Res_101	88.2 (0.4)	70.8 (0.6)	69.2 (0.8)	79.7 (0.2)
DenseNet	89.4 (0.5)	71.3 (0.8)	70.6 (0.6)	79.3 (0.3)

Ragusa, E., Cambria, E., Zunino, R., & Gastaldo, P. (2019). A survey on deep learning in image polarity detection: Balancing generalization performances and computational costs. *Electronics*, 8(7), 783.



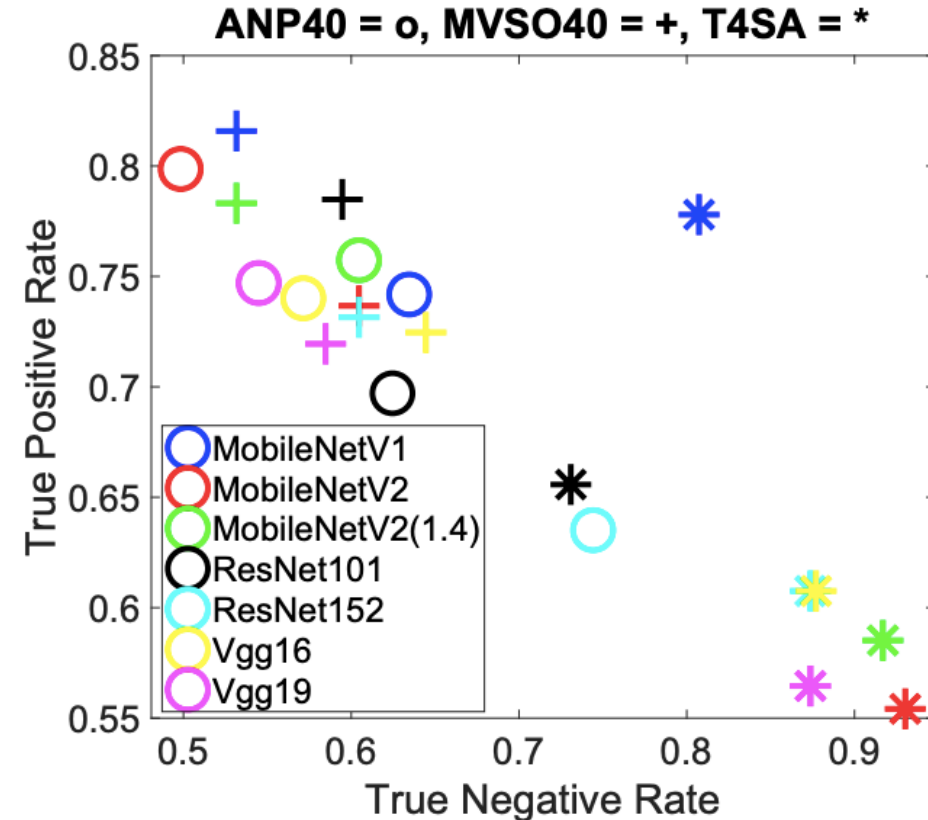
MobileNets for Image Polarity detection

- Mobile nets
 - Key-point: Depth-wise separable convolutions
 - Pros
 - Very high trade-off accuracy/compute cost
 - Explicitly designed for embedded systems
 - Contra
 - Lower accuracy for standard benchmark
 - Object classification
 - Object detection

Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Le, Q. V. (2019). Searching for mobilenetv3. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1314-1324).



MobileNets for Image Polarity detection

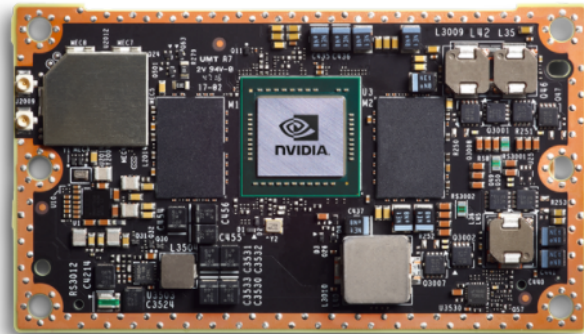


Ragusa, E., Gianoglio, C., Zunino, R., & Gastaldo, P. (2020). Image Polarity Detection on Resource-Constrained Devices. *IEEE Intelligent Systems*.



Edge Accelerators

Jetson TX2



Movidius NCS





Implementation

Architecture	Movidius			Jetson		
	Latency (ms)	Memory (GB)	Power (Watt)	Latency (ms)	Memory (GB)	Power (Watt)
MobileNet_v1	43.98	0.18	1.17	12.0	0.93	1.65/5.76
MobileNet_v2	43.40	0.17	0.87	16.3	2.56	2.04/6.13
MobileNet_v2(1.4)	61.59	0.20	1.20	24.2	3.69	2.12/6.37
Res_101	405.49	0.93	1.42	22.2	4.22	1.56/7.55
Res_152	607.26	1.23	1.56	31.7	8.42	1.55/7.71
Vgg_16	864.84	2.68	-	43.7	11.60	2.41/7.94
Vgg_19	1046.84	2.79	-	58.0	11.59	2.38/7.59

Ragusa, E., Gianoglio, C., Zunino, R., & Gastaldo, P. (2020). Image Polarity Detection on Resource-Constrained Devices. *IEEE Intelligent Systems*.



Role of saliency



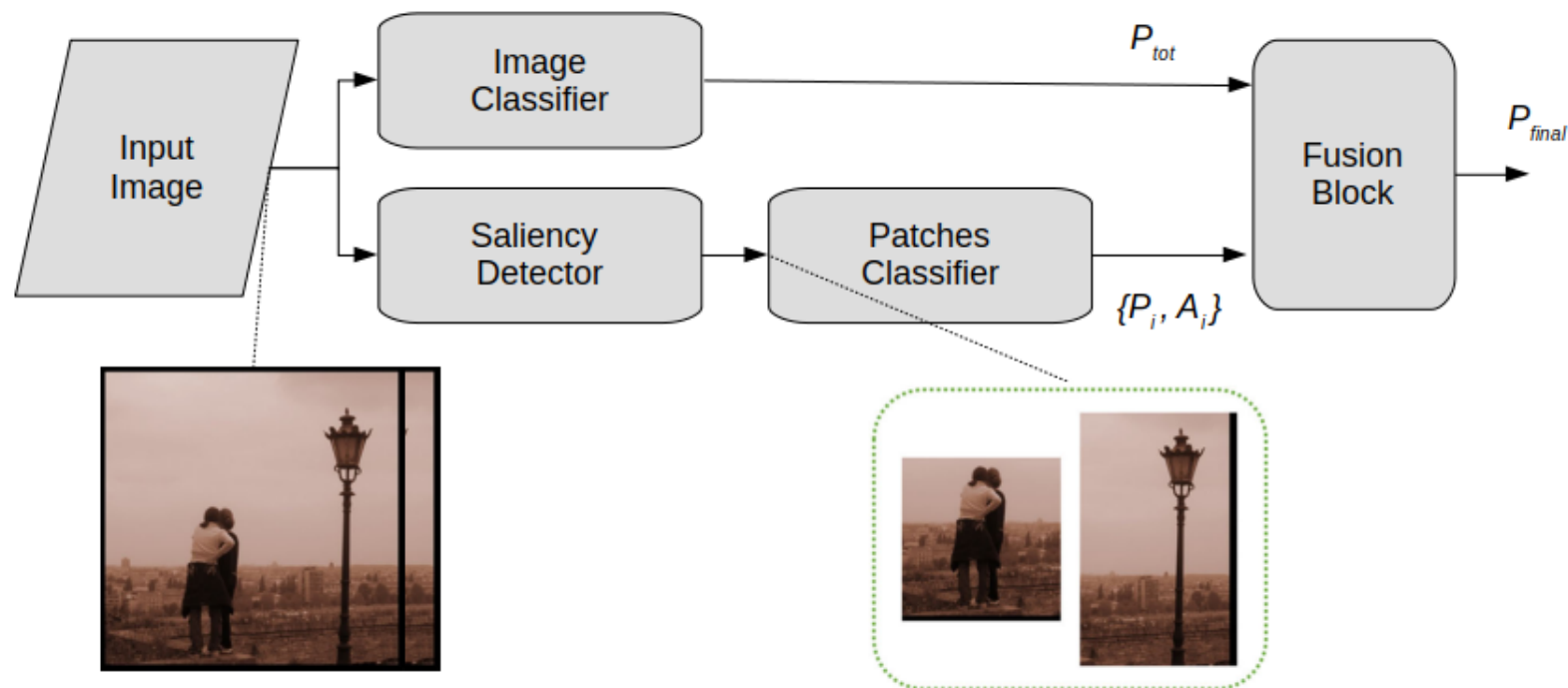


Role of saliency





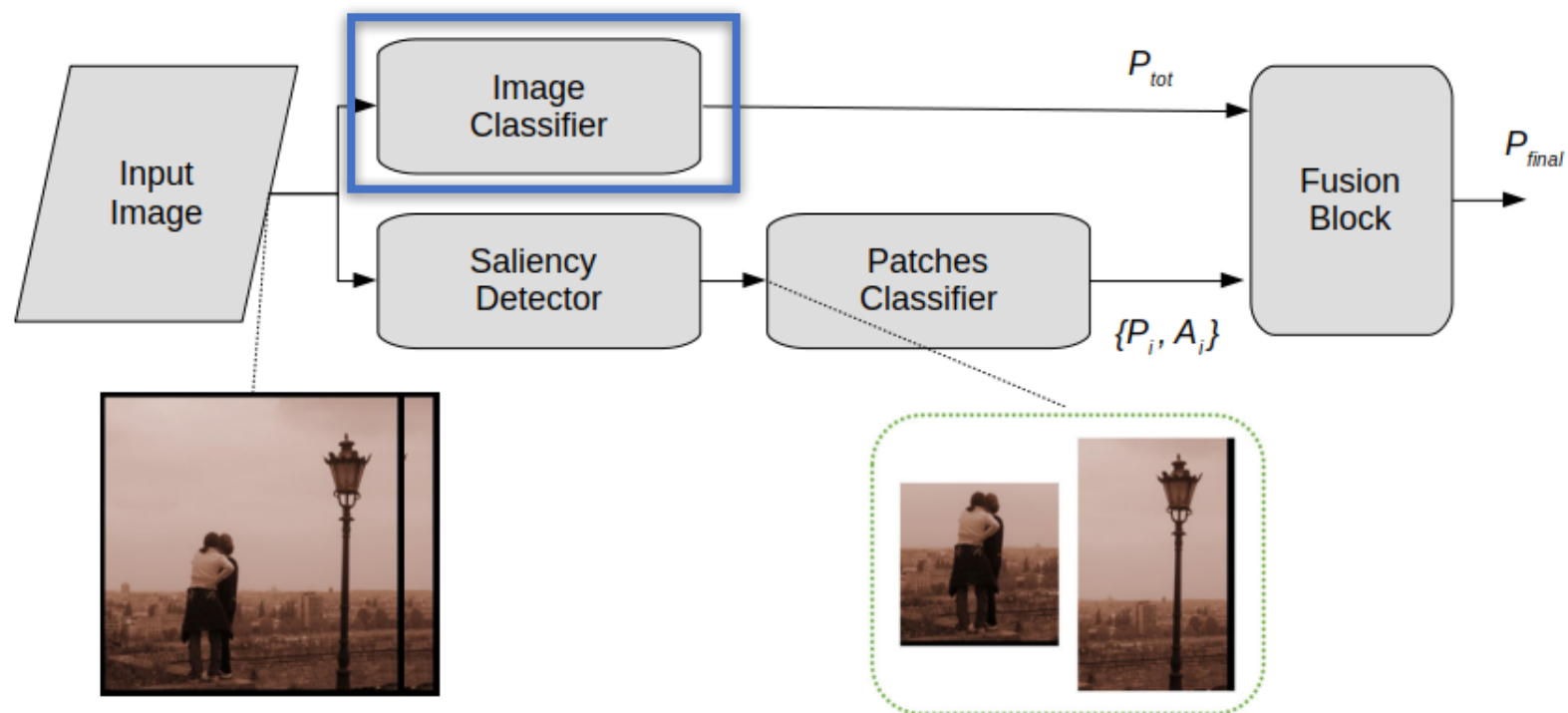
Proposal



Ragusa, E., Apicella, T., Gianoglio, C., Zunino, R., & Gastaldo, P. (2020). An hardware-aware image polarity detector enhanced with visual attention. *IJCNN 2020, accepted*

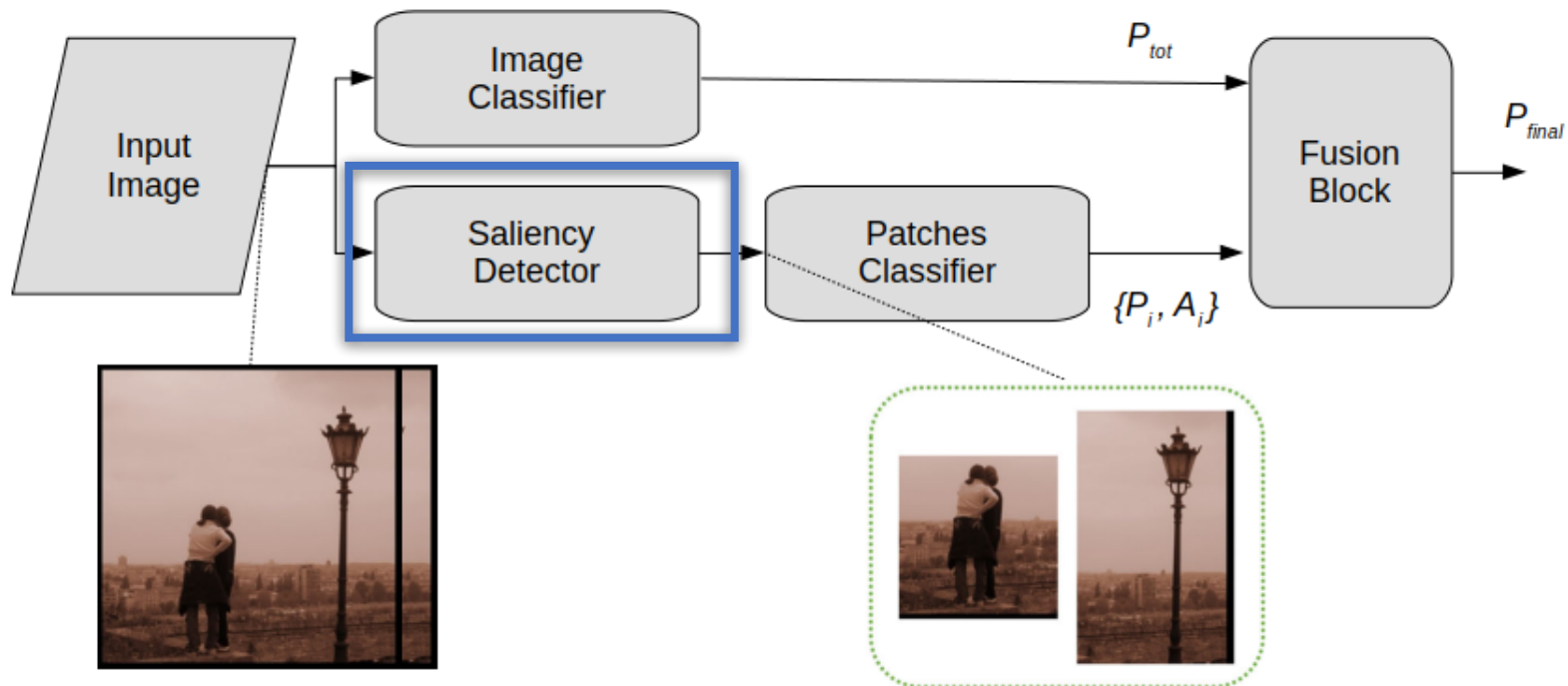


Proposal



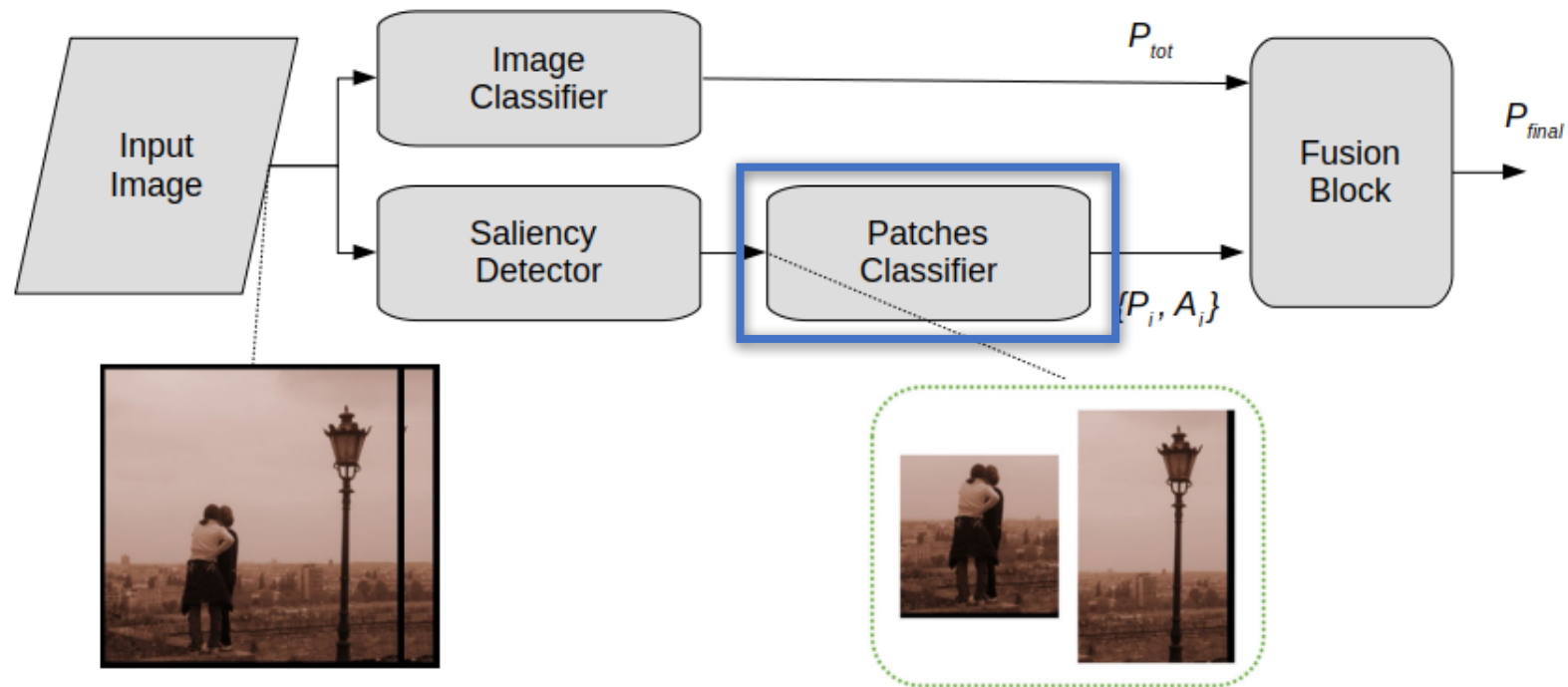


Proposal



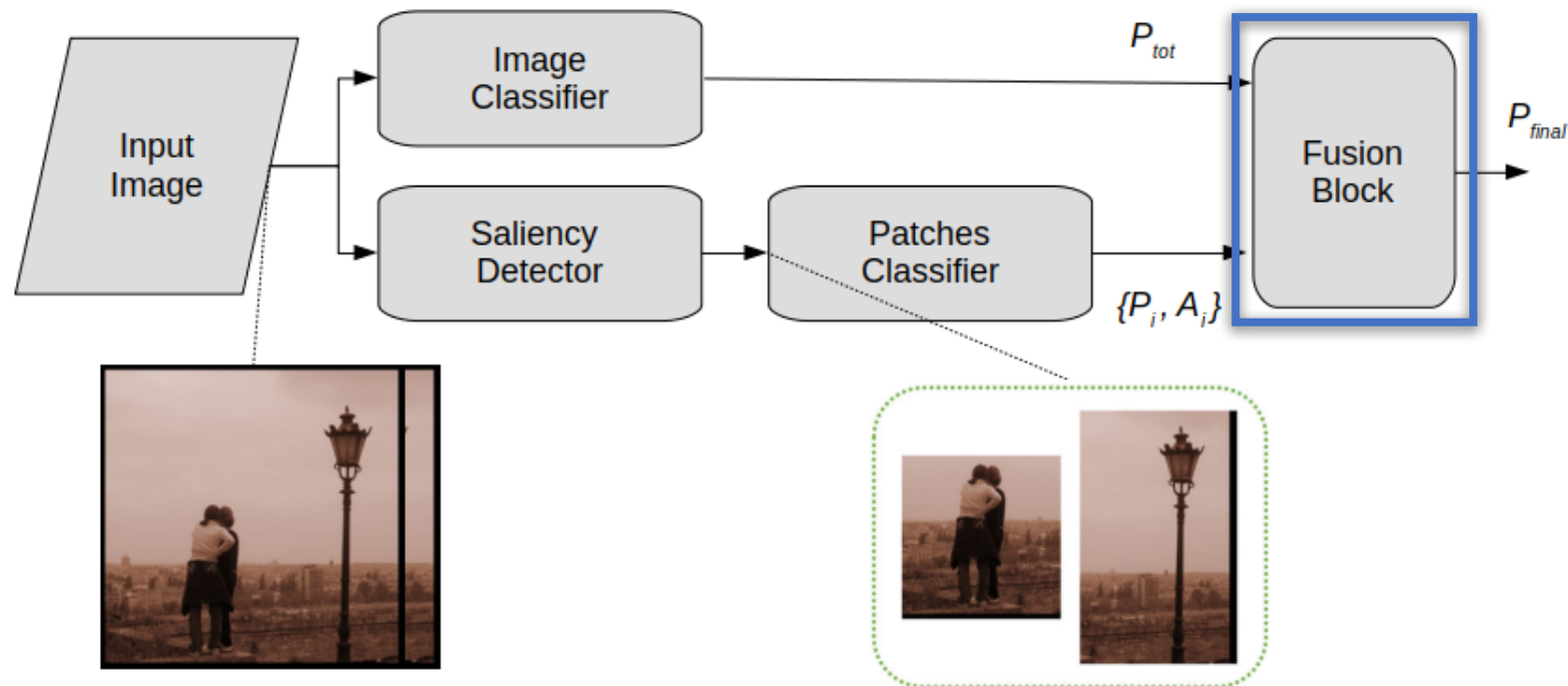


Proposal





Proposal





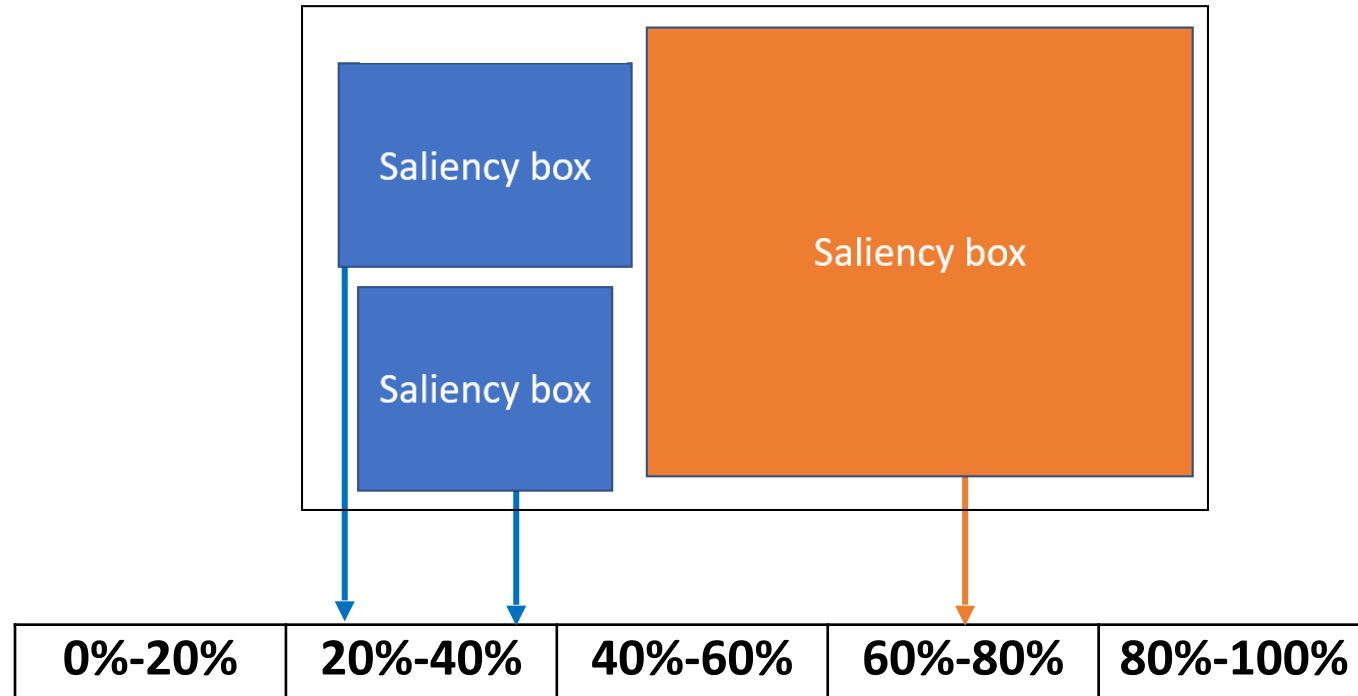
Blocks

- Image classifier and Patches classifier
 - MobileNetV1
- Saliency detector
 - SSD-MobileNetV1
- Fusion block
 - Rule based

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., ... & Murphy, K. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7310-7311).



Fusion Block



Experiments

- Saliency detection
- Proposal
- Deployment on smartphones





Saliency detection

- Network
 - SSD MobileNetV1
 - TensorFlow Object Detect API
 - Iterations 50000
- ILSVRC-2014
 - 127030 images
- SOS
 - 3951 images



Results

- Standard hold out:
 - 90% training
 - 10% test

TABLE I
SALIENCY DETECTOR PERFORMANCES

Saliency detector	True boxes	Predicted	IoU > 75%
SSD_MobileNet_ILSVRC	13982	13118	10207
SSD_MobileNet_SOS	13982	14723	9272

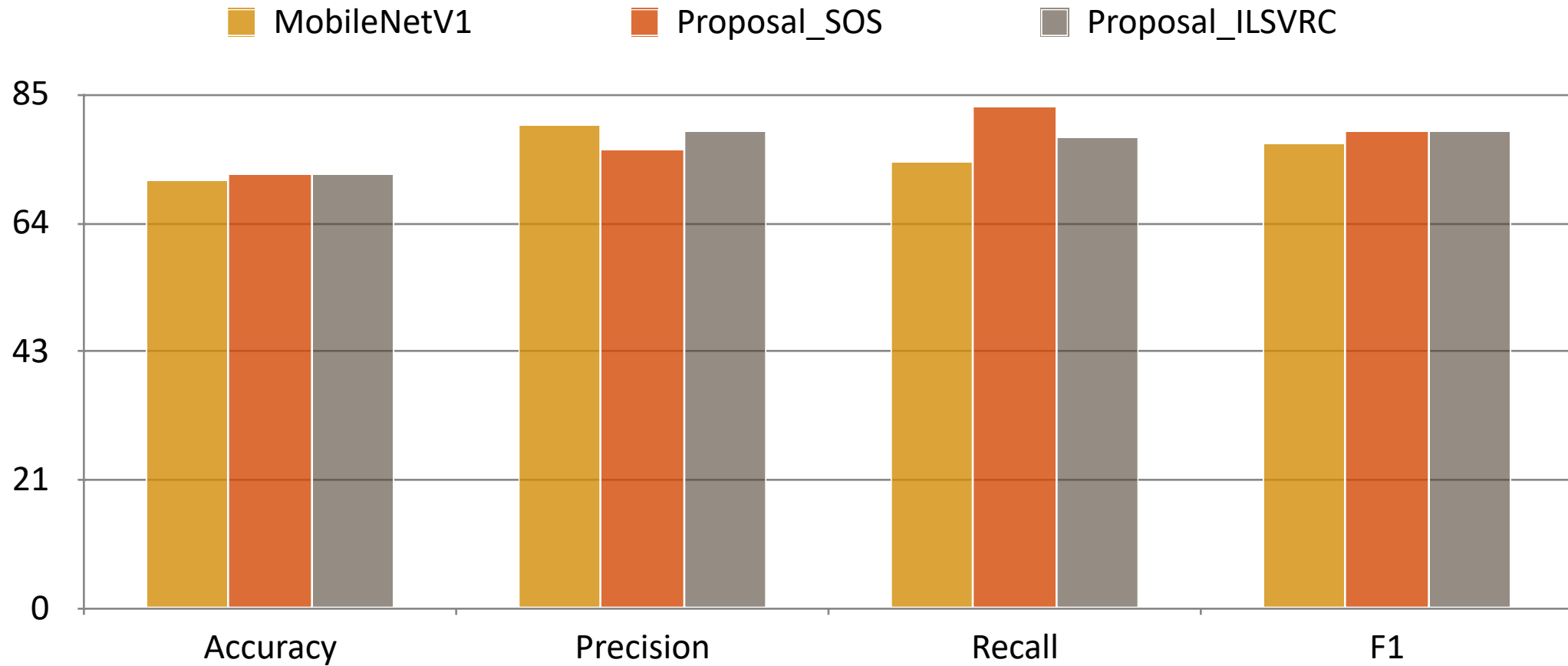


Experiments

- MobileNetV1
 - TensorFlow-Slim library
- Dataset
 - Training:
 - ANP40
 - Test:
 - Twitter
 - MVSA_eq
 - MVSA_maj

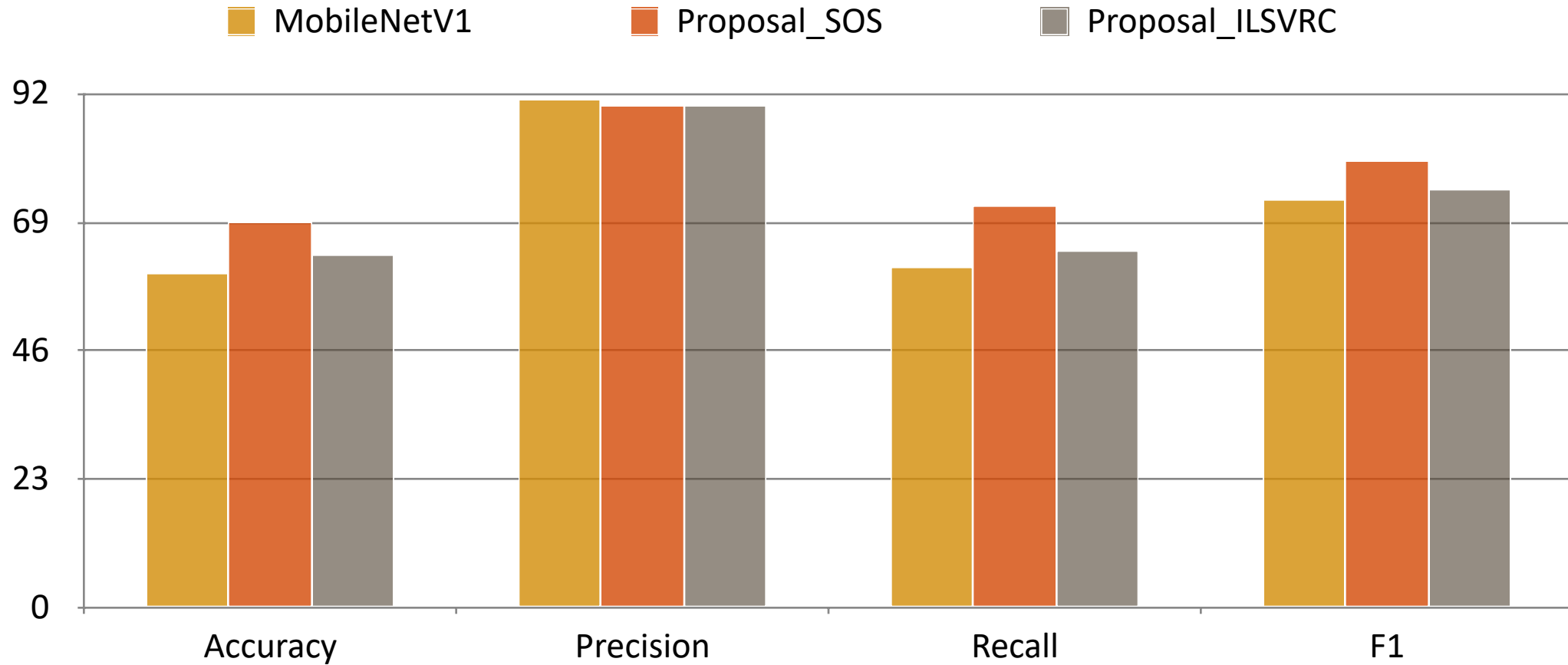


Experiments: Twitter



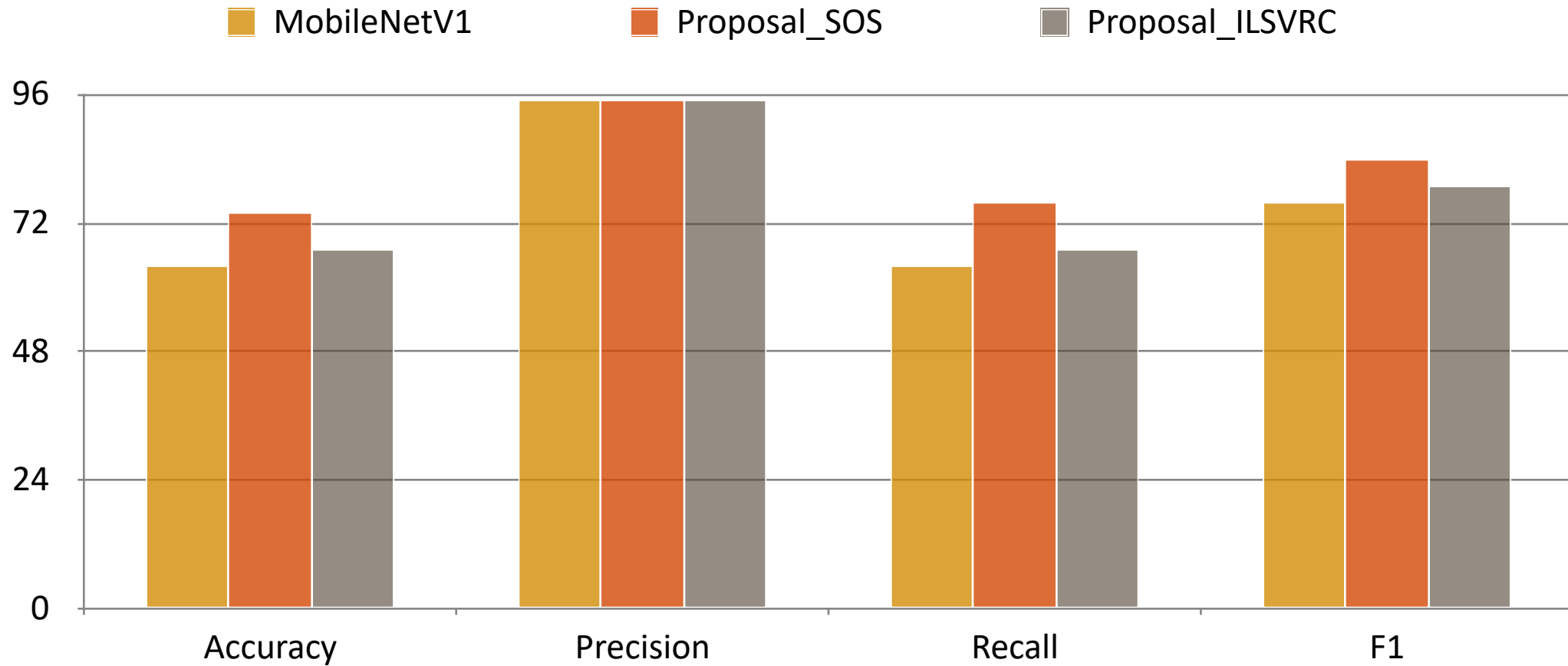


Experiments: MVSA_maj





Experiments: MVSA_eq





Smartphones Implementation

- 5 android smartphones
 - Only processor
 - No gpu or Npu
- Inference time for an image:
 - 0.5 to 0.9 seconds FP32
 - 0.4 to 0.8 second FP16



Conclusion

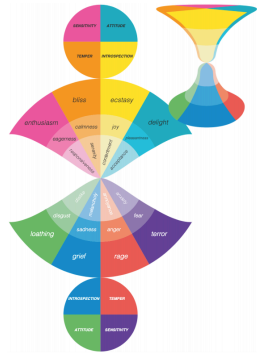
- This part of the speech presented:
 - Strategies, algorithms and hardware devices for image polarity detection on embedded devices
 - Result regarding accuracy and hardware performances.



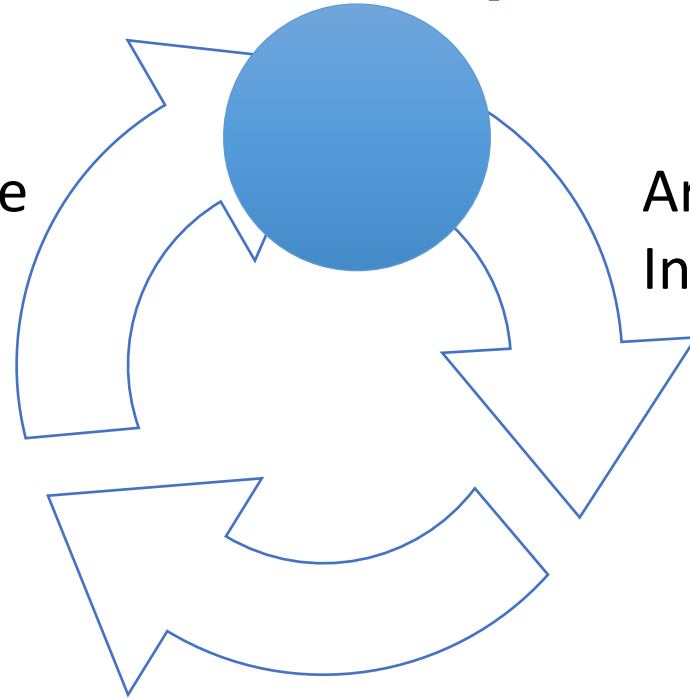
Cognitive models and computational resources



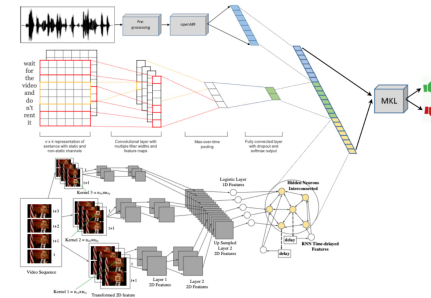
Hardware-algorithm loop



Cognitive models



Artificial Intelligence



Hardware Resources



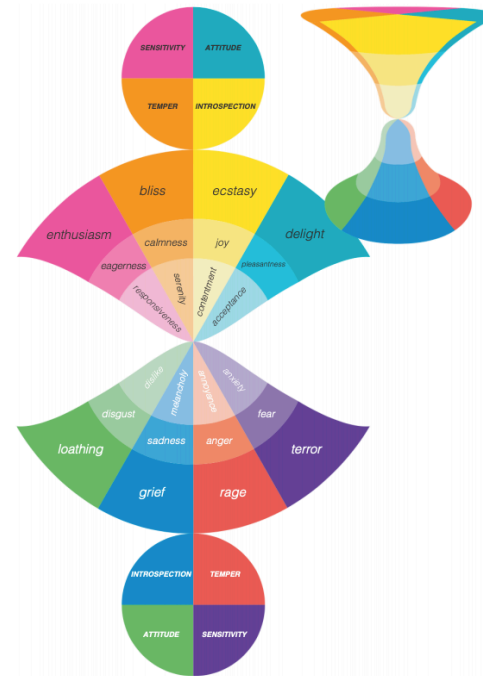
Cognitive models and computational resources

- Goal
 - Study the coherence between
 - Cognitive model
 - Hourglass of emotions
 - Computational resource
 - Affective space
- Output
 - Experimental protocol
 - Analysis

Ragusa, E., Gastaldo, P., Zunino, R., Ferrarotti, M. J., Rocchia, W., & Decherchi, S. (2019). Cognitive Insights into Sentic Spaces Using Principal Paths. *Cognitive Computation*, 11(5), 656-675.



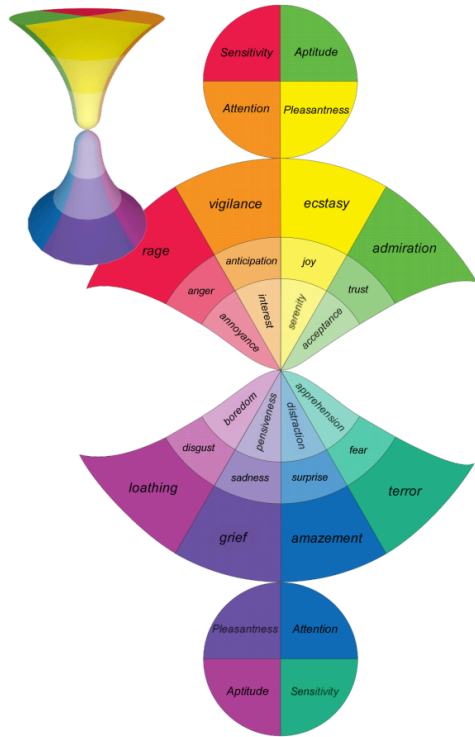
Cognitive model: The Hourglass of Emotions



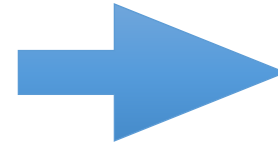
Susanto, Y., Livingstone, A., Ng, B. C., & Cambria, E. (2020). The hourglass model revisited. *IEEE Intelligent Systems*, 35(5).



Cognitive model: The Hourglass of Emotions



State of mind



4 dimensional vector

Cambria, E., Livingstone, A., & Hussain, A. (2012). The hourglass of emotions. In *Cognitive behavioural systems* (pp. 144-157). Springer, Berlin, Heidelberg.



Motivation

- Affective space is 300 dimensional
 - Inspection is difficult
- Standard measures cosine/euclidean does not represent completely cognitive models
- Dimensionality reduction techniques could be helpful for visualization but limit the insight about the relative position of the data in the 300- dimensional affective space



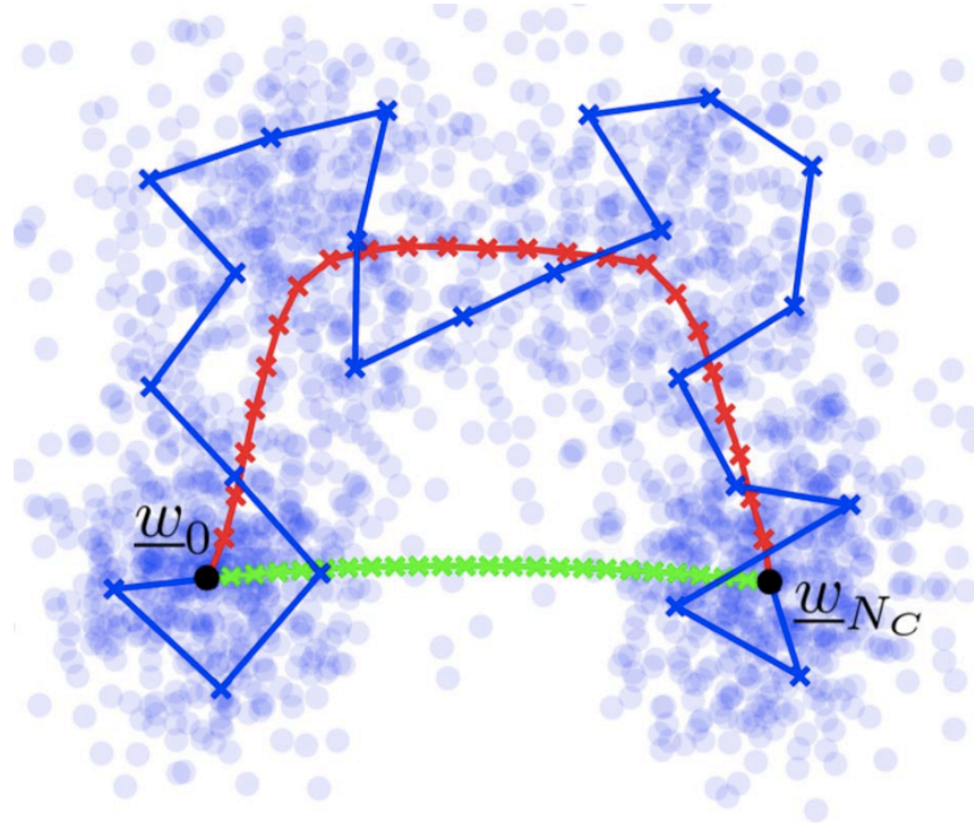
Principal path algorithm

$$\min_{\mathbf{W}} \frac{\gamma}{2} \sum_{i=1}^N \sum_{j=1}^{N_c} \|\mathbf{x}_i - \mathbf{w}_j\|^2 \delta(u_i, j) + \frac{\lambda}{2} \sum_{i=0}^{N_c} \|\mathbf{w}_{i+1} - \mathbf{w}_i\|^2$$

Ferrarotti, M. J., Rocchia, W., & Decherchi, S. (2018). Finding principal paths in data space. *IEEE transactions on neural networks and learning systems*, 30(8), 2449-2462.



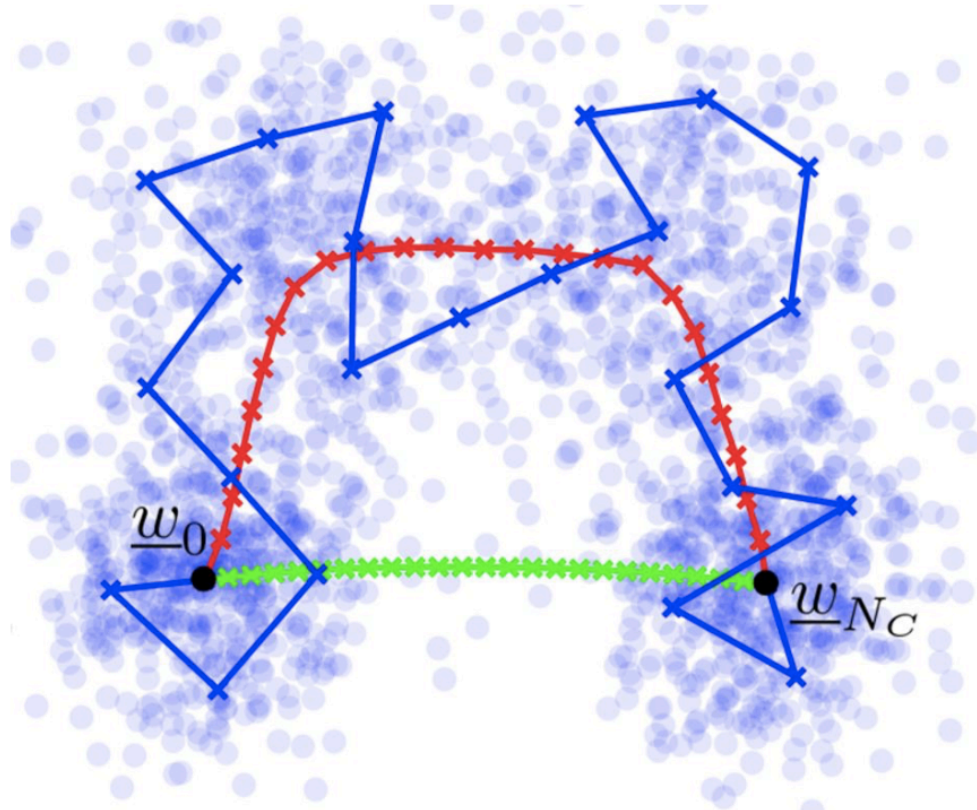
Principal path algorithm



Ferrarotti, M. J., Rocchia, W., & Decherchi, S. (2018). Finding principal paths in data space. *IEEE transactions on neural networks and learning systems*, 30(8), 2449-2462.



Principal path algorithm

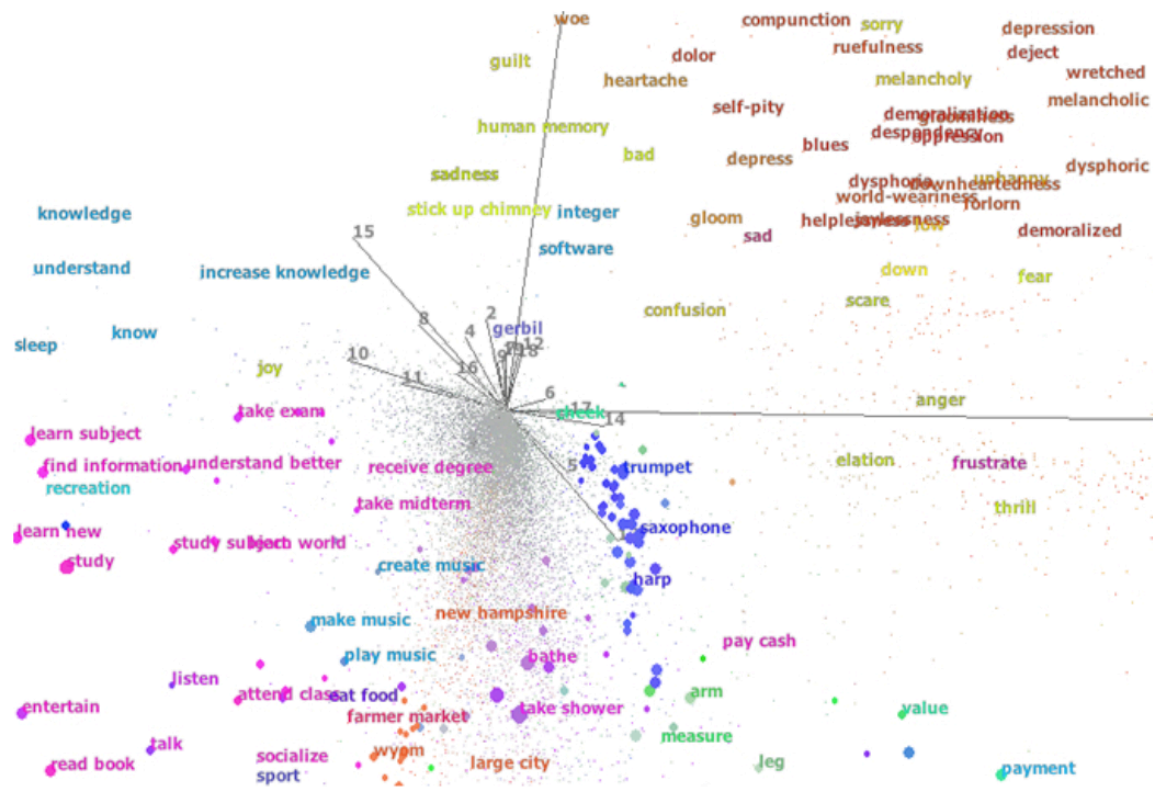
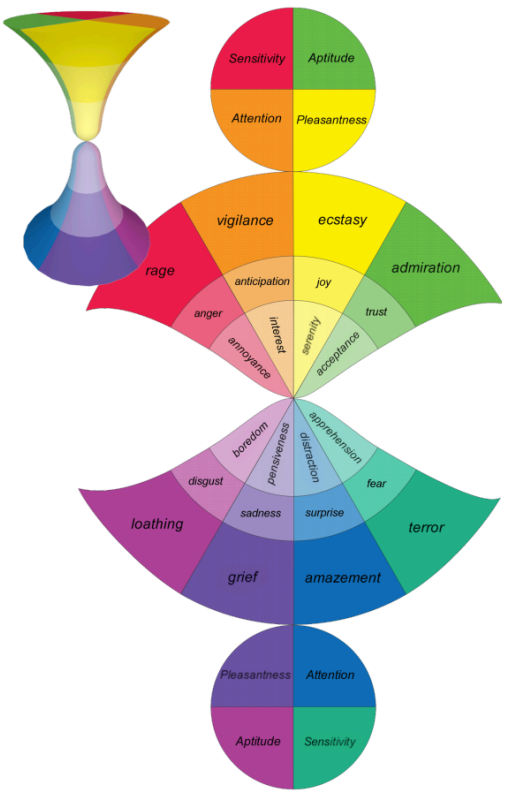


Given two points, it describes data manifolds using a ordered set of points.

Ferrarotti, M. J., Rocchia, W., & Decherchi, S. (2018). Finding principal paths in data space. *IEEE transactions on neural networks and learning systems*, 30(8), 2449-2462.



Paths in AffectiveSpace



Ragusa, E., Gastaldo, P., Zunino, R., Ferrarotti, M. J., Rocchia, W., & Decherchi, S. (2019). Cognitive Insights into Sentic Spaces Using Principal Paths. *Cognitive Computation*, 11(5), 656-675.

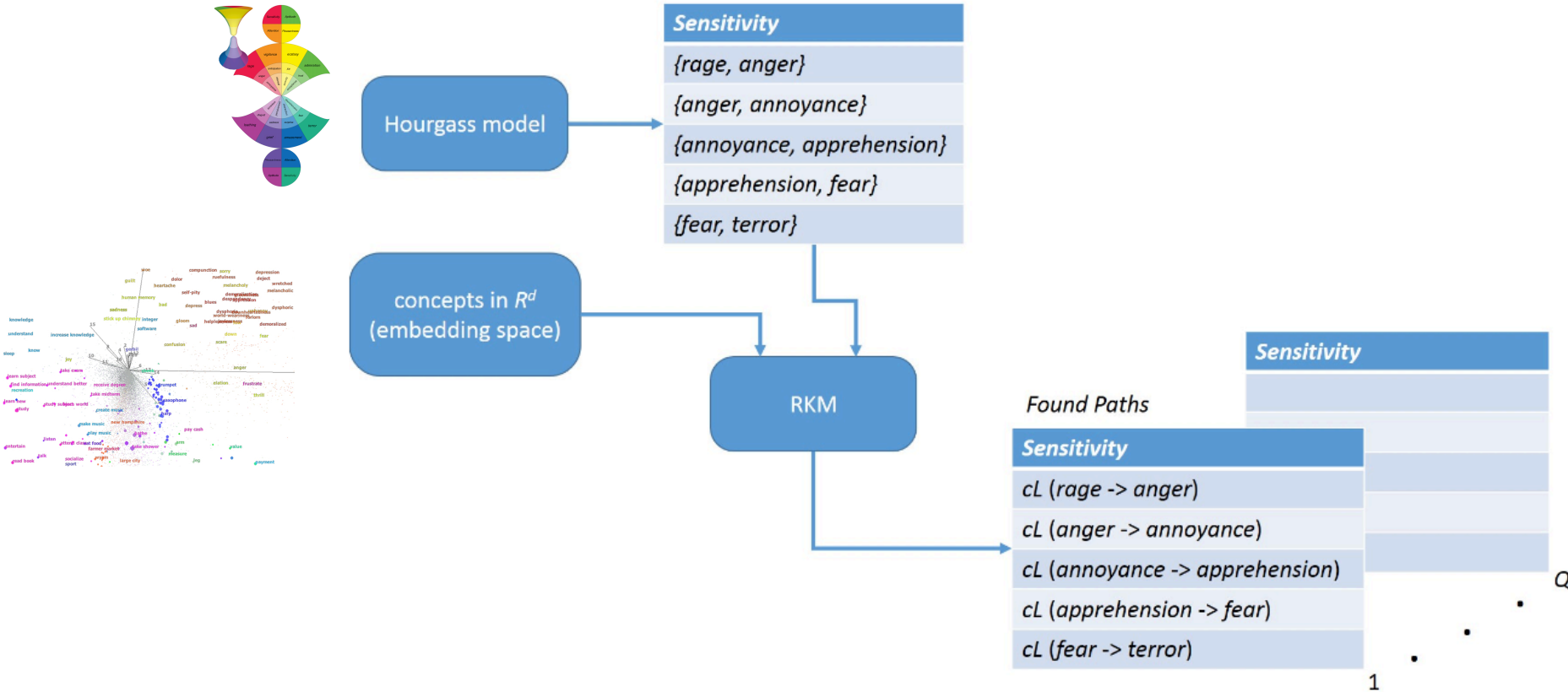


Paths in AffectiveSpace

- Procedure
 - Hourglass of emotions:
 - 4 dimensions
 - 6 landmarks concept for each dimension
 - Study data manifolds that connects landmarks concepts in Affective Space.

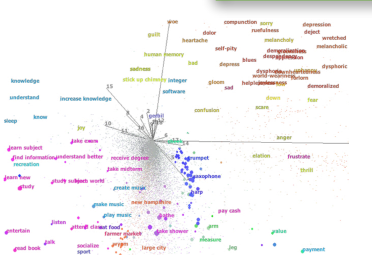
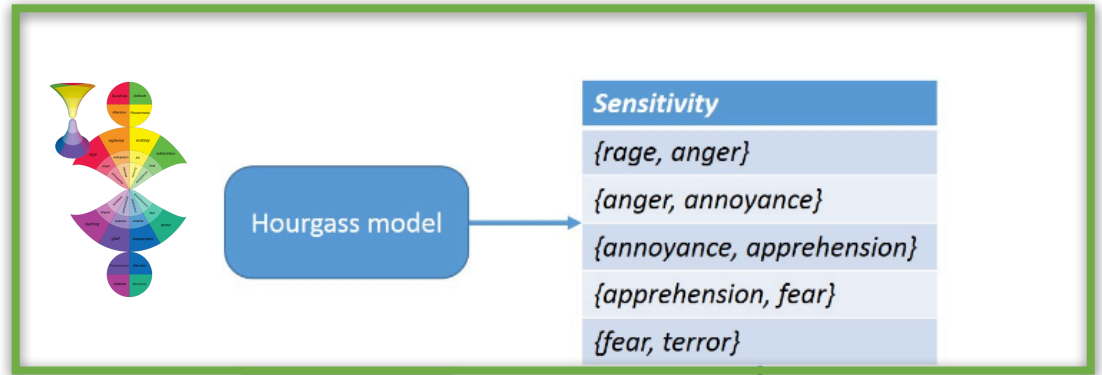


Protocol 1



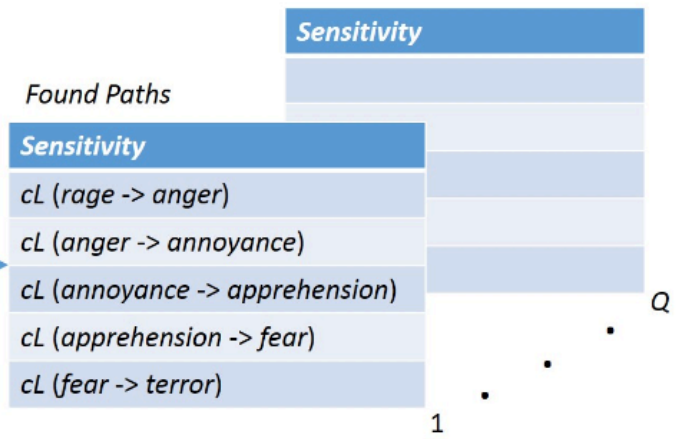


Protocol 1



concepts in R^d
(embedding space)

RKM



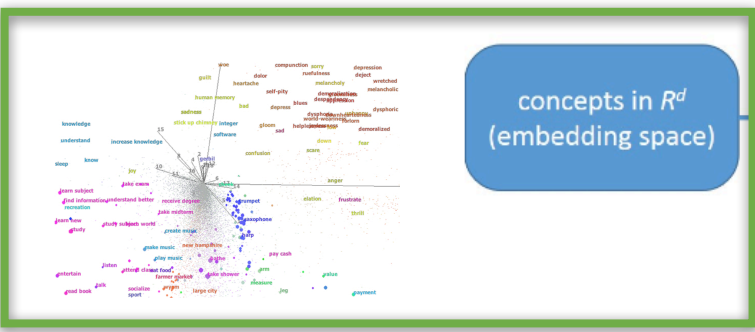


Protocol 1



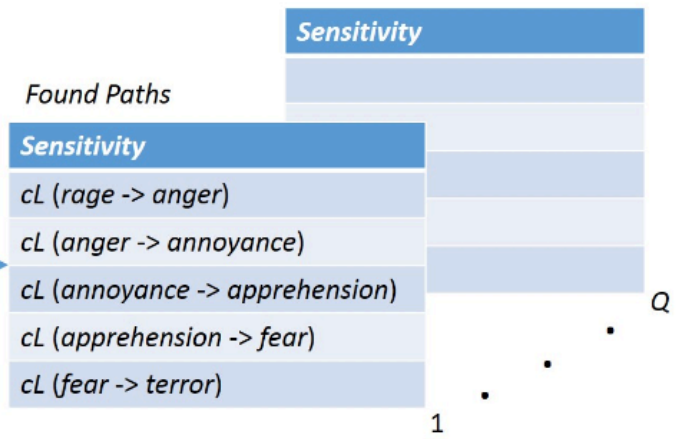
Hourglass model

Sensitivity
{rage, anger}
{anger, annoyance}
{annoyance, apprehension}
{apprehension, fear}
{fear, terror}



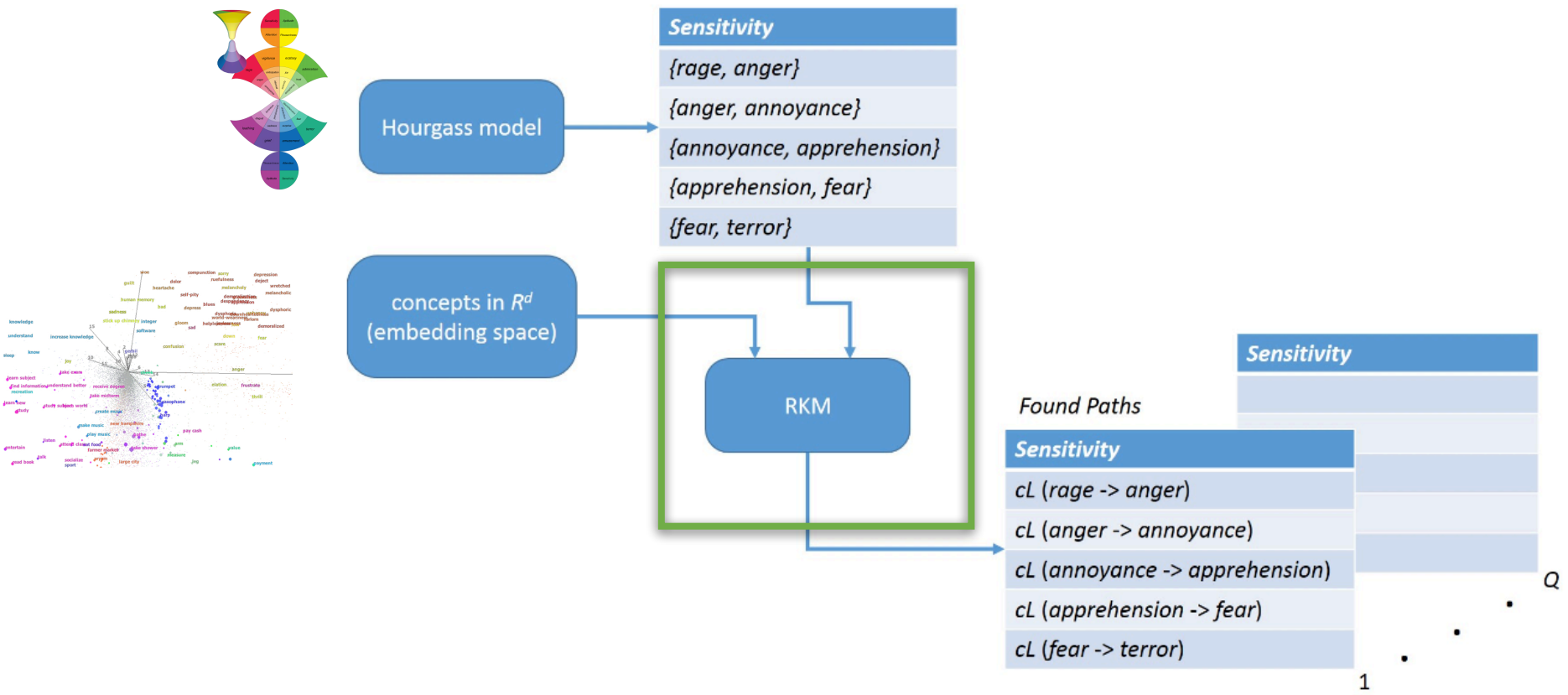
concepts in R^d (embedding space)

RKM



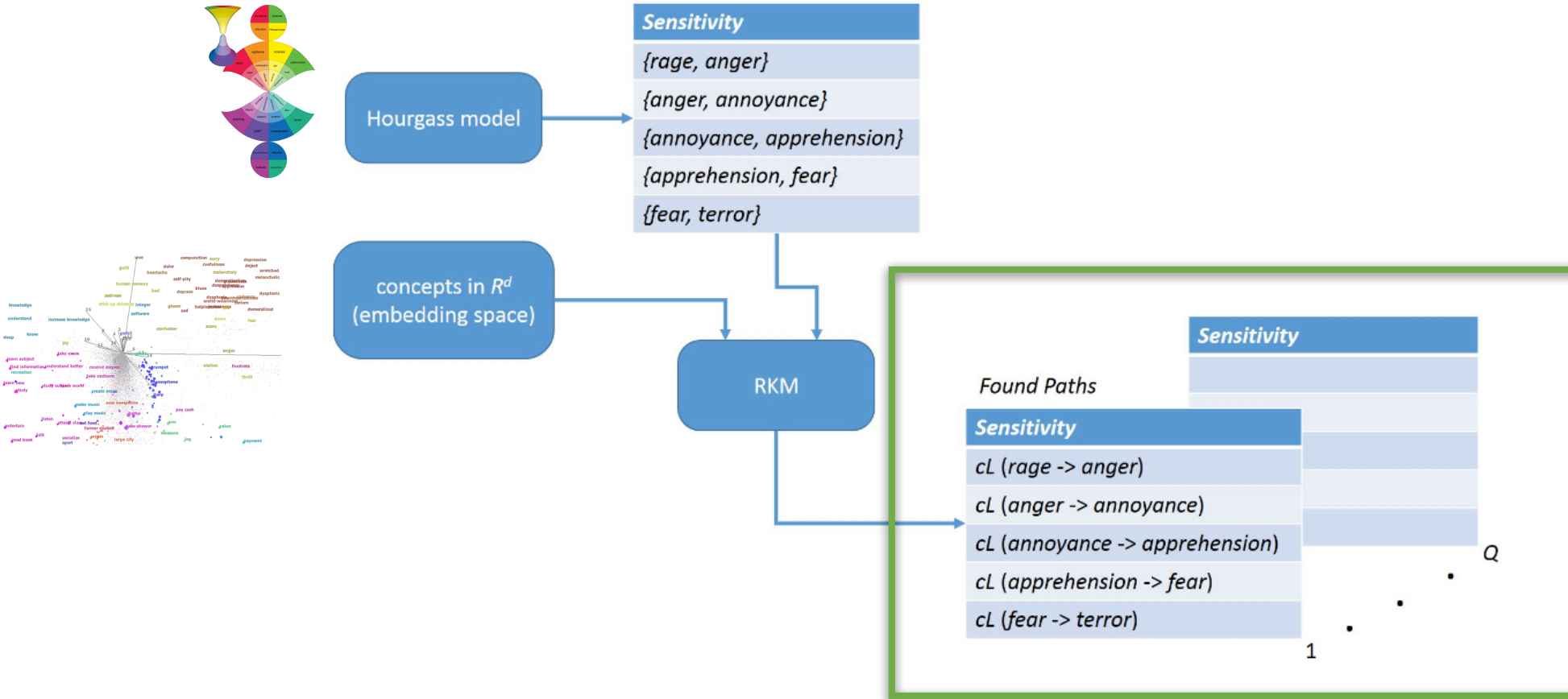


Protocol 1



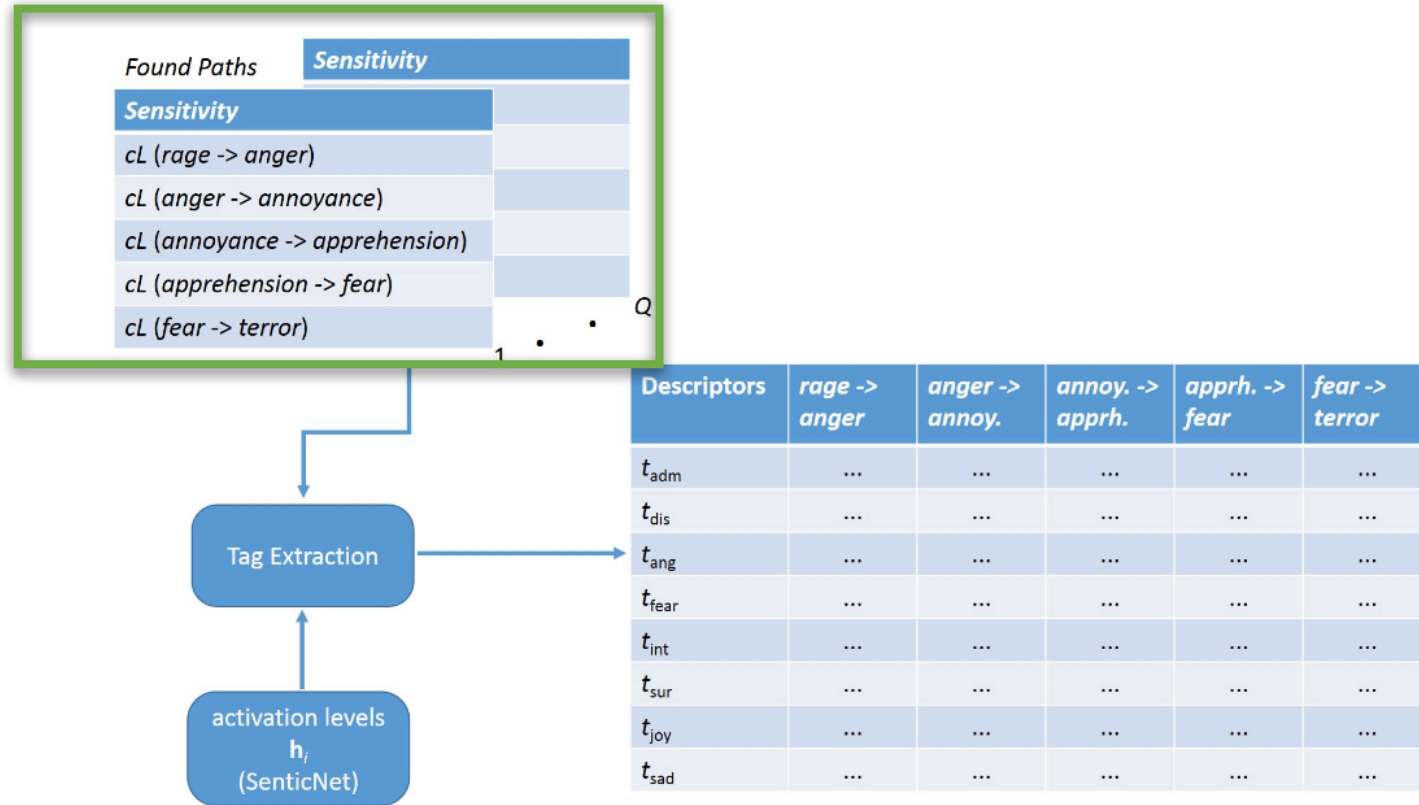


Protocol 1



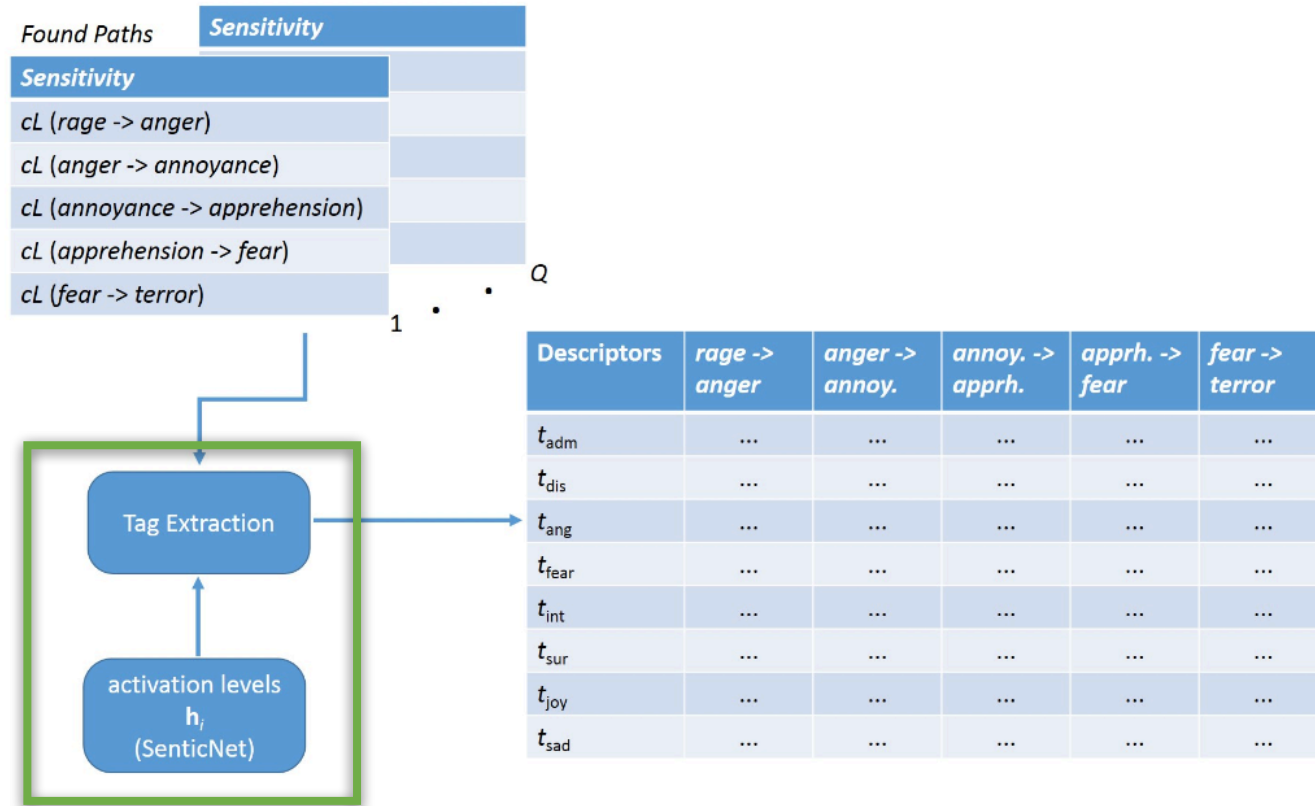


Protocol 2



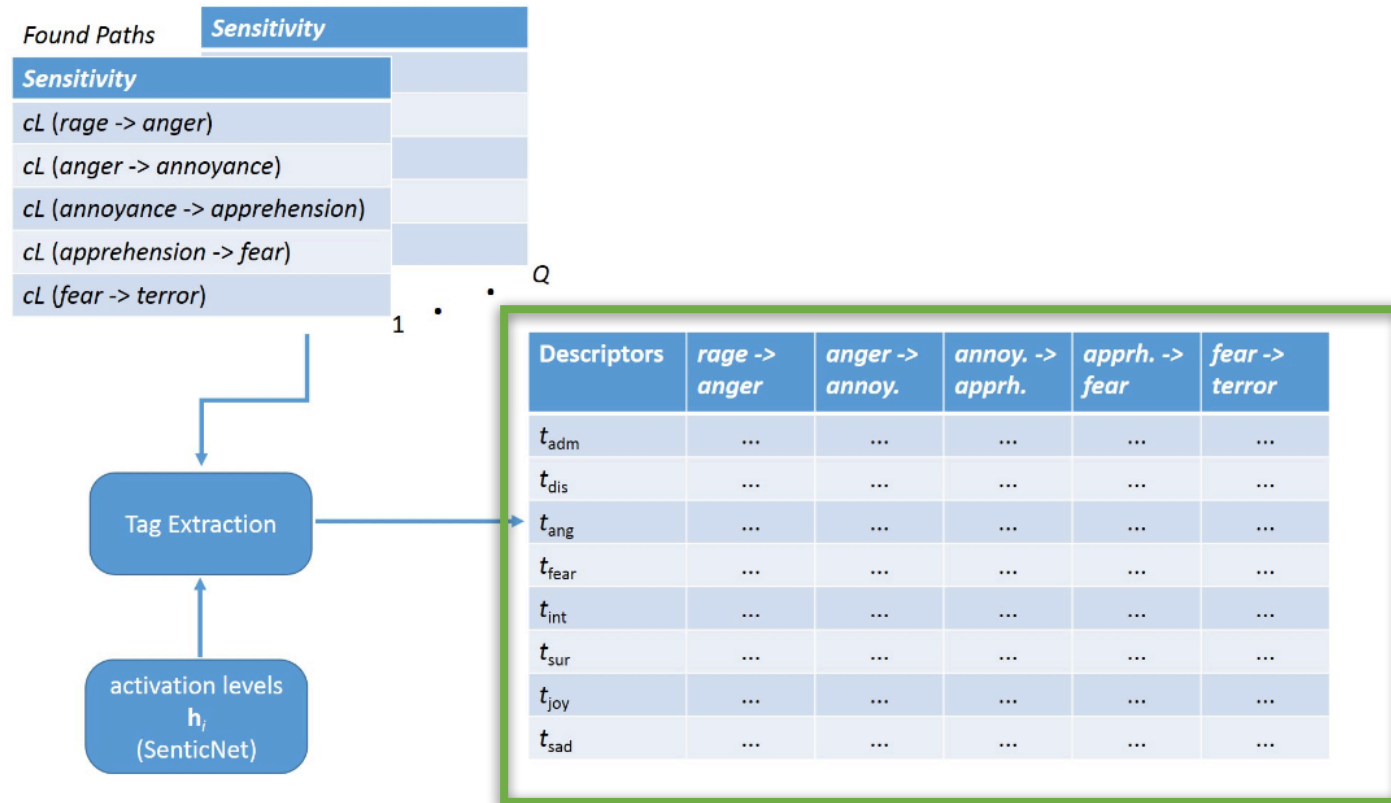


Protocol 2



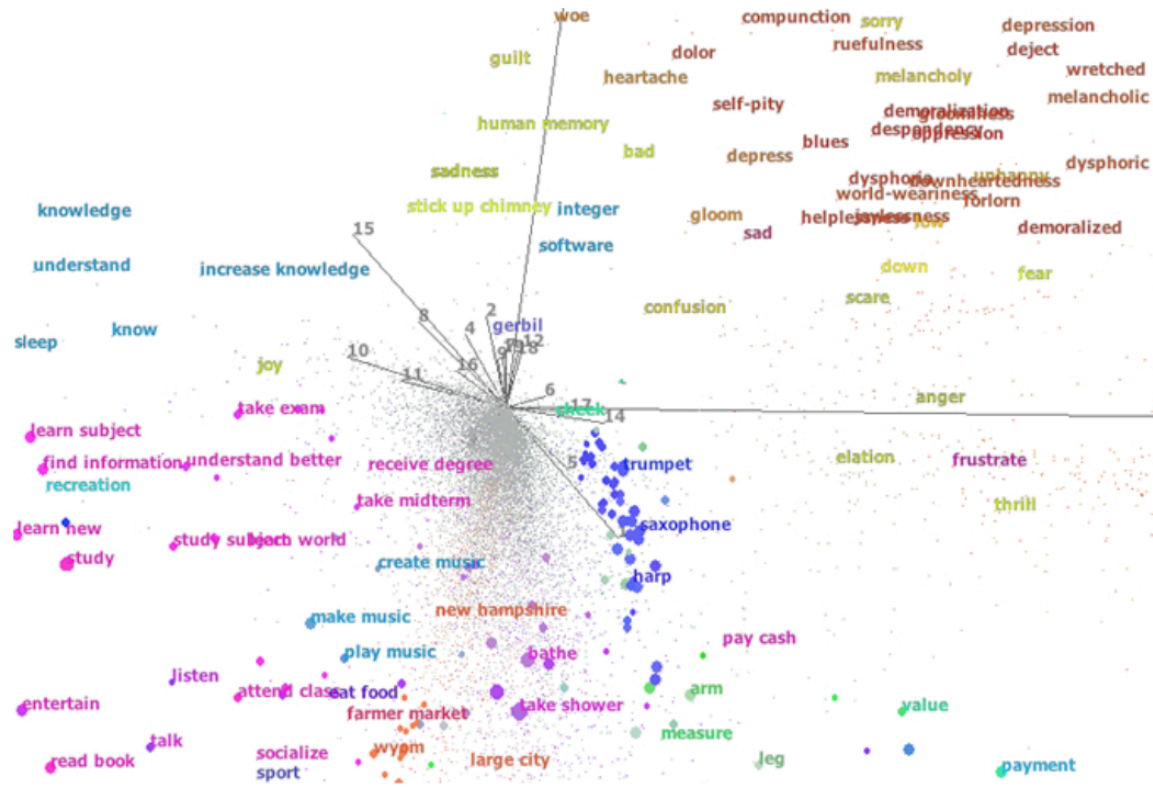
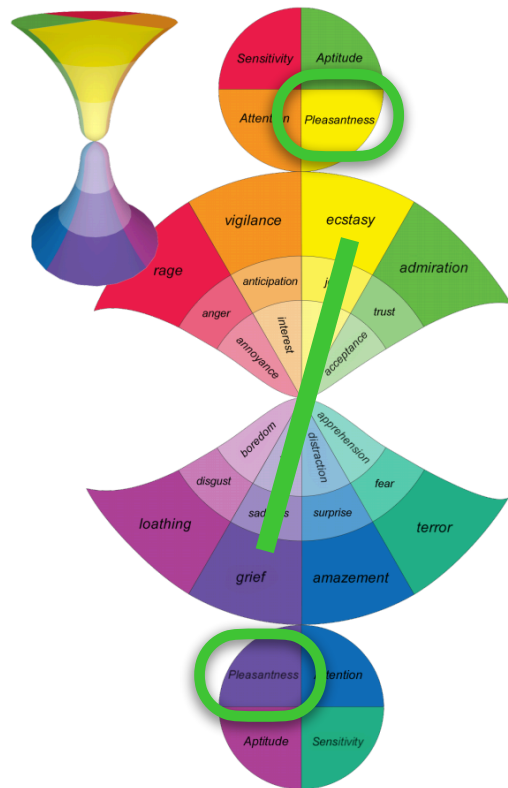


Protocol 2



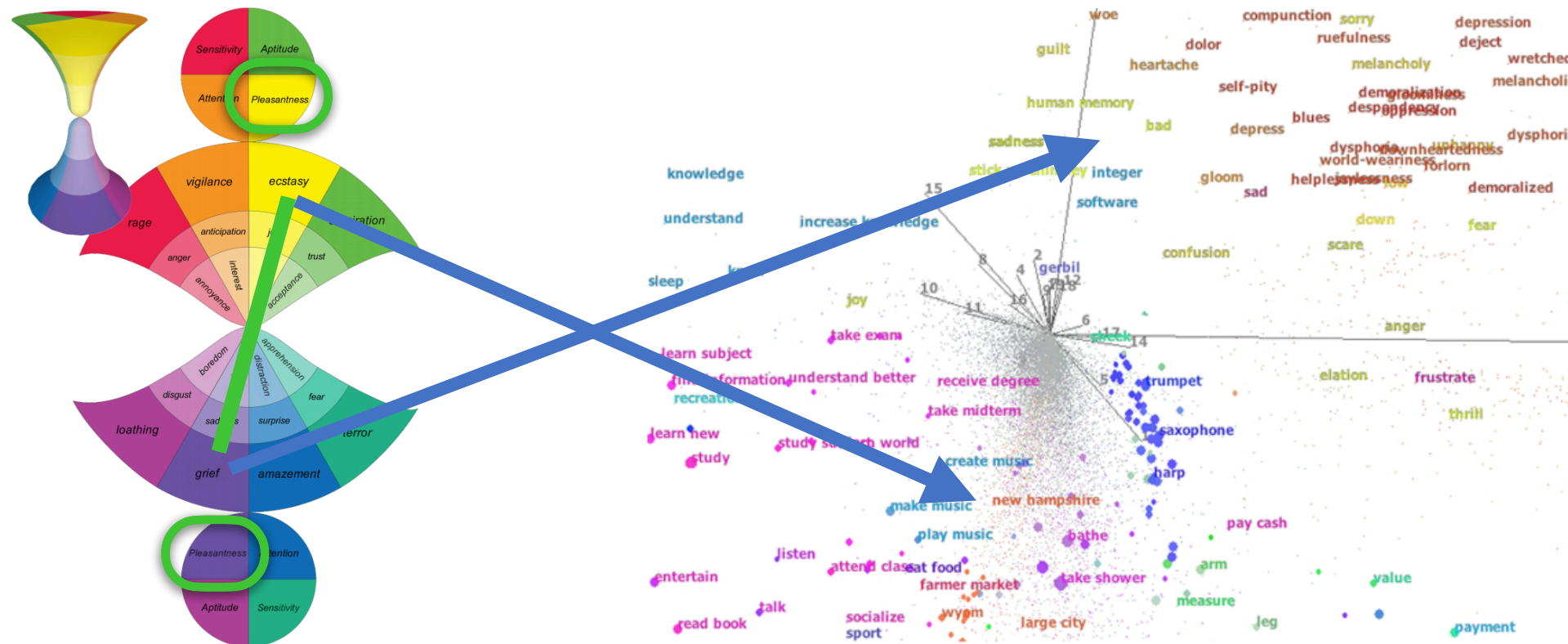


Cognitive model: The Hourglass of Emotions



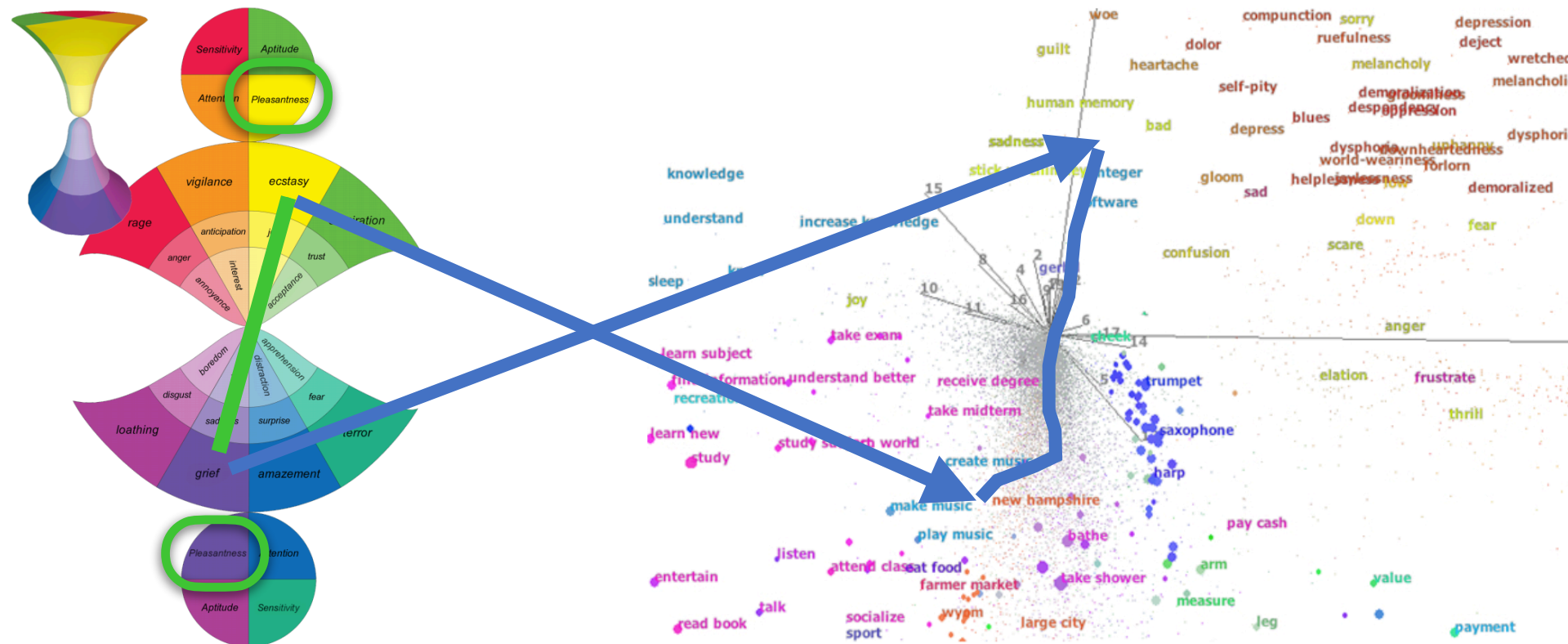


Cognitive model: The Hourglass of Emotions



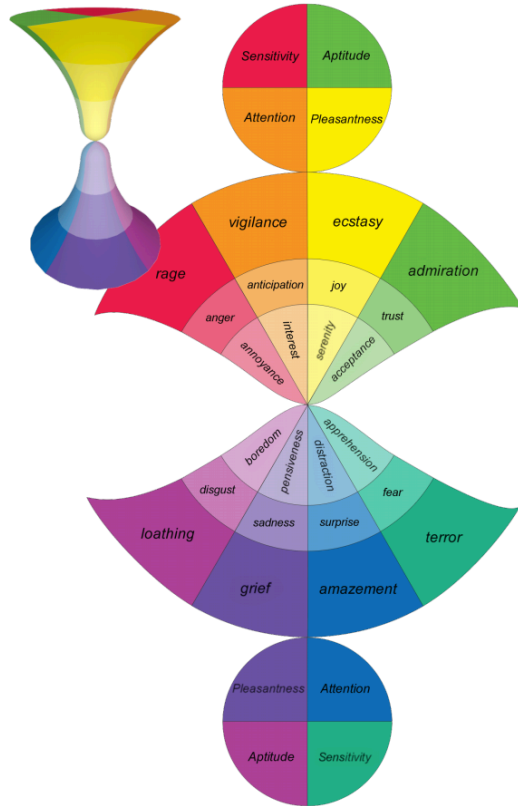


Cognitive model: The Hourglass of Emotions

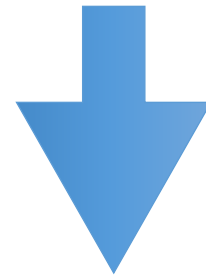




Tags



Centroids in 300 dimensions



SenticNet

Points in 4 dimensions

Sensitivity: -0.9

Aptitude: 0.7

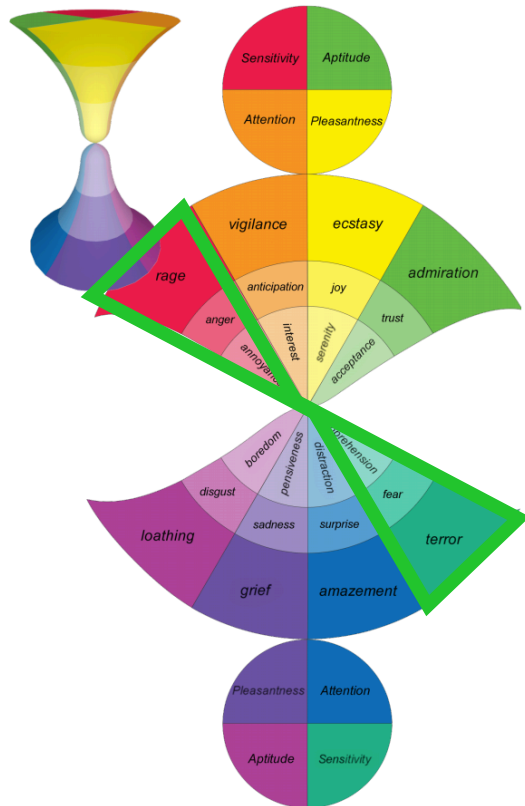
Pleasantness: 0.3

Attention: -0.4

Cambria, E., Li, Y., Xing, F. Z., Poria, S., & Kwok, K. (2020). Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. CIKM.



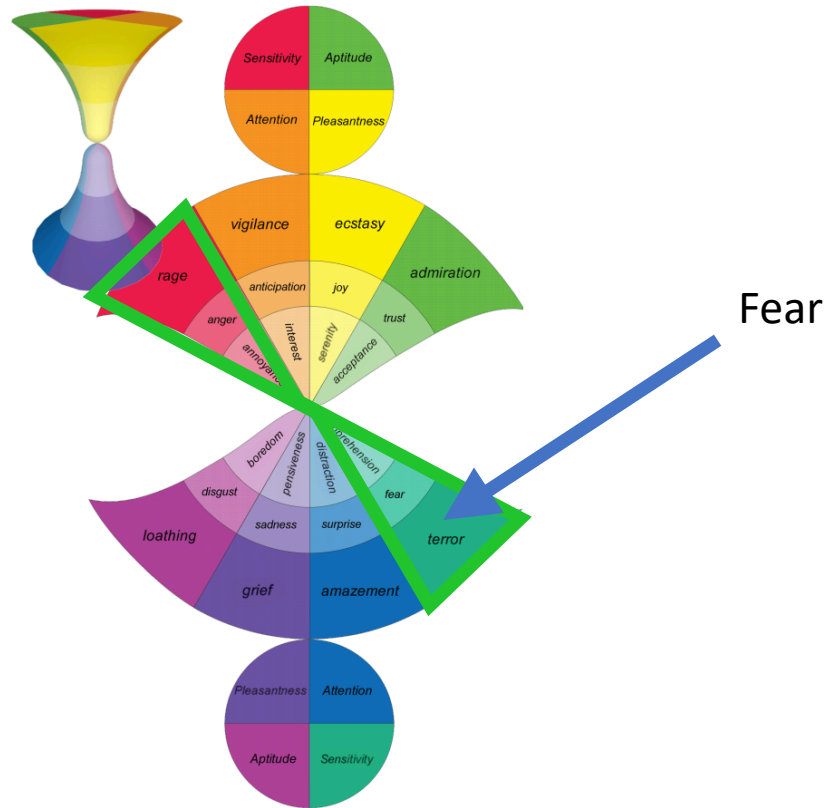
Tags



Concept:
Sensitivity: -0.9
Aptitude: 0.7
Pleasantness: 0.3
Attention: -0.4

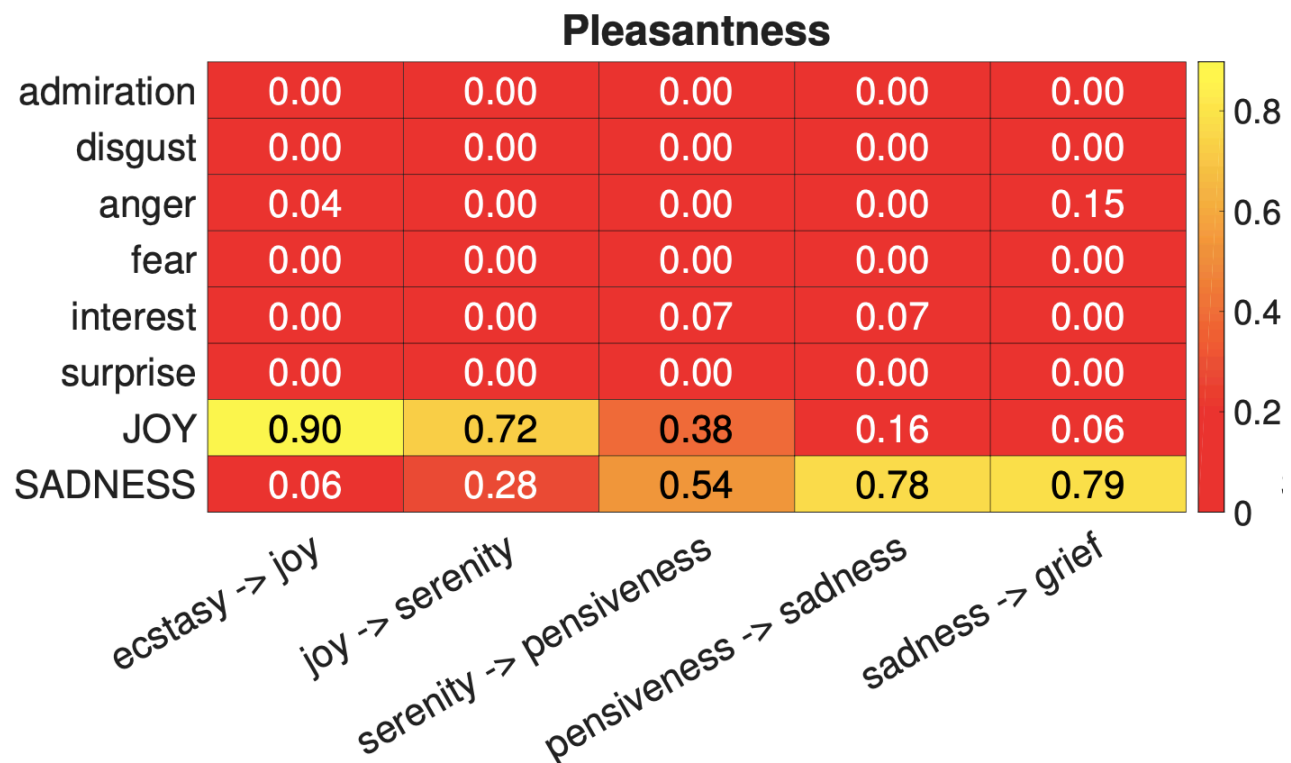


Tags



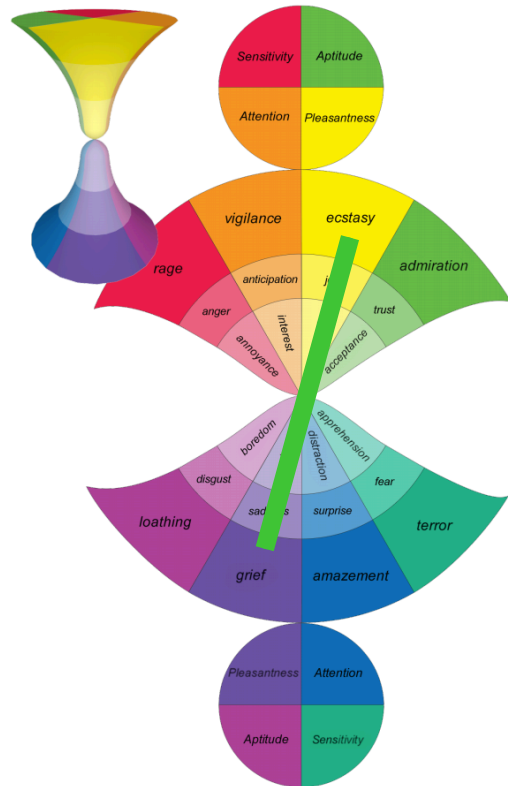
Concept:
Sensitivity: -0.9
Aptitude: 0.7
Pleasantness: 0.3
Attention: -0.4

Results





Results



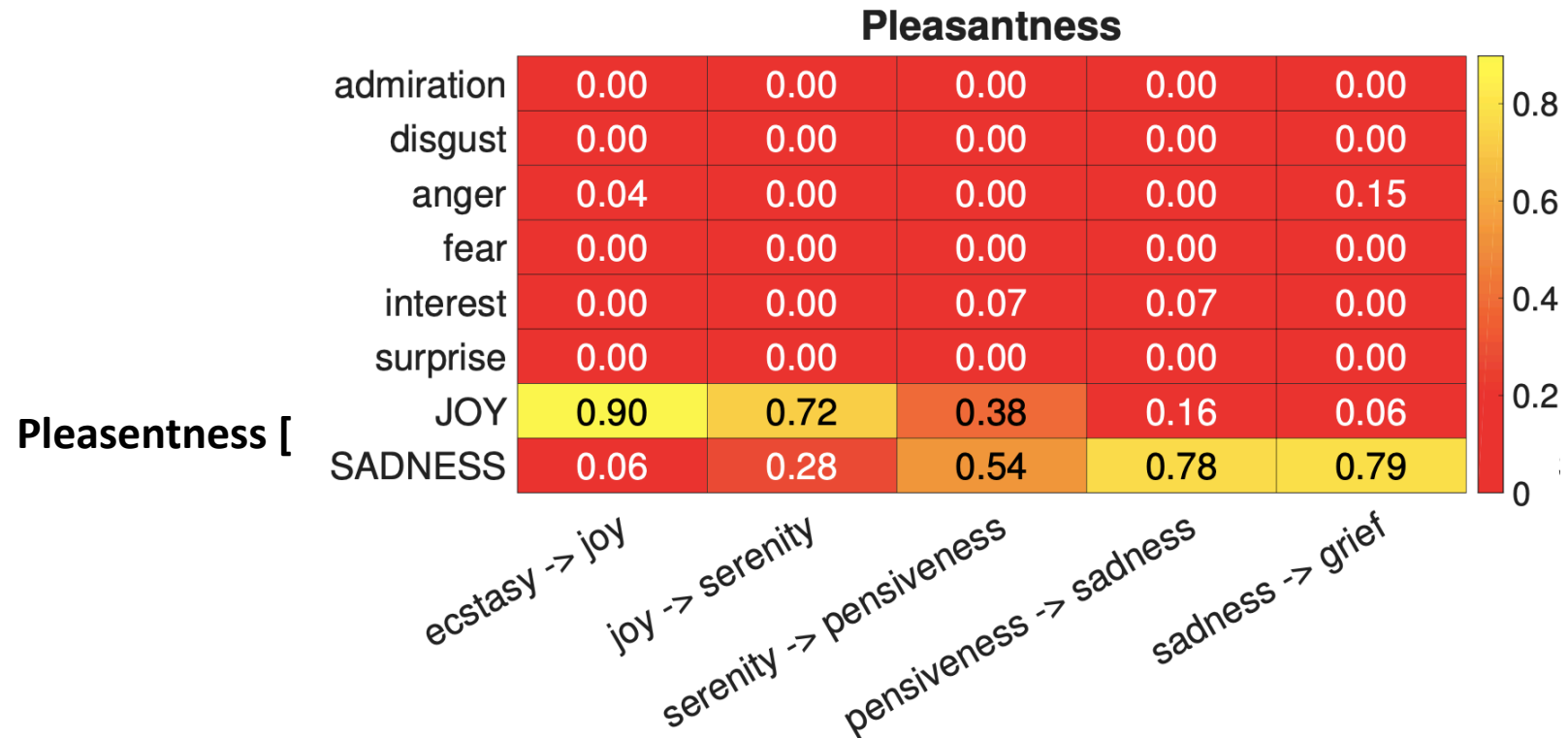
Pleasantness

ration	0.00	0.00	0.00	0.00	0.00
isgust	0.00	0.00	0.00	0.00	0.00
anger	0.04	0.00	0.00	0.00	0.15
fear	0.00	0.00	0.00	0.00	0.00
terest	0.00	0.00	0.07	0.07	0.00
rprise	0.00	0.00	0.00	0.00	0.00
JOY	0.90	0.72	0.38	0.16	0.06
NESS	0.06	0.28	0.54	0.78	0.79

ecstasy -> joy
 joy -> serenity
 serenity -> pensiveness
 pensiveness -> sadness
 sadness -> grief



Results





Conclusion

- Hardware aware solutions needs domain knowledges:
 - Good models
 - AI algorithms
 - Hardware and software resources
- Examples
 - Image polarity detection
 - Embeddings analysis



Thank you for your attention