Assignment 2

11/26/2020

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Saliency4ASD: Visual attention modeling for Autism Spectrum Disorder

Huiyu Duan

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Contents

1 Background

2 ICME2019 Grand Challenge

3 Solutions and Results

4 Assignment Requirements

Contents

1 Background

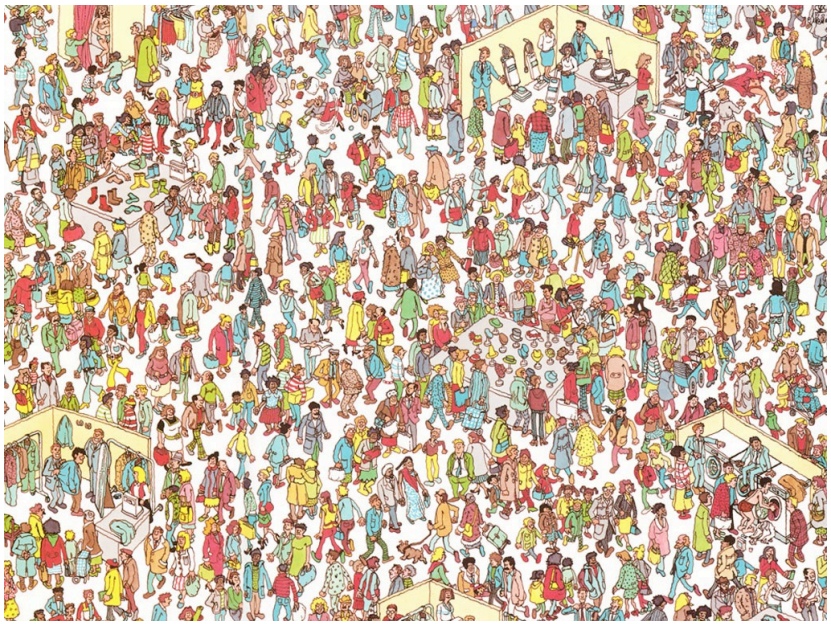
2 ICME2019 Grand Challenge

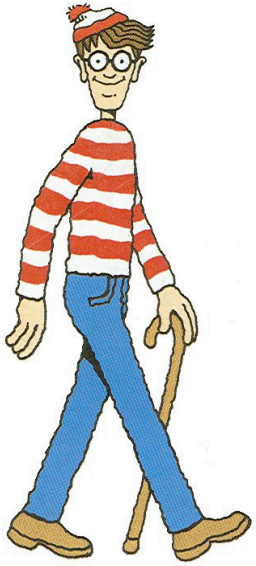
3 Solutions and Results

4 Assignment Requirements

Introduction to visual attention

- Natural visual scenes are cluttered and contain many different objects that cannot all be processed simultaneously.

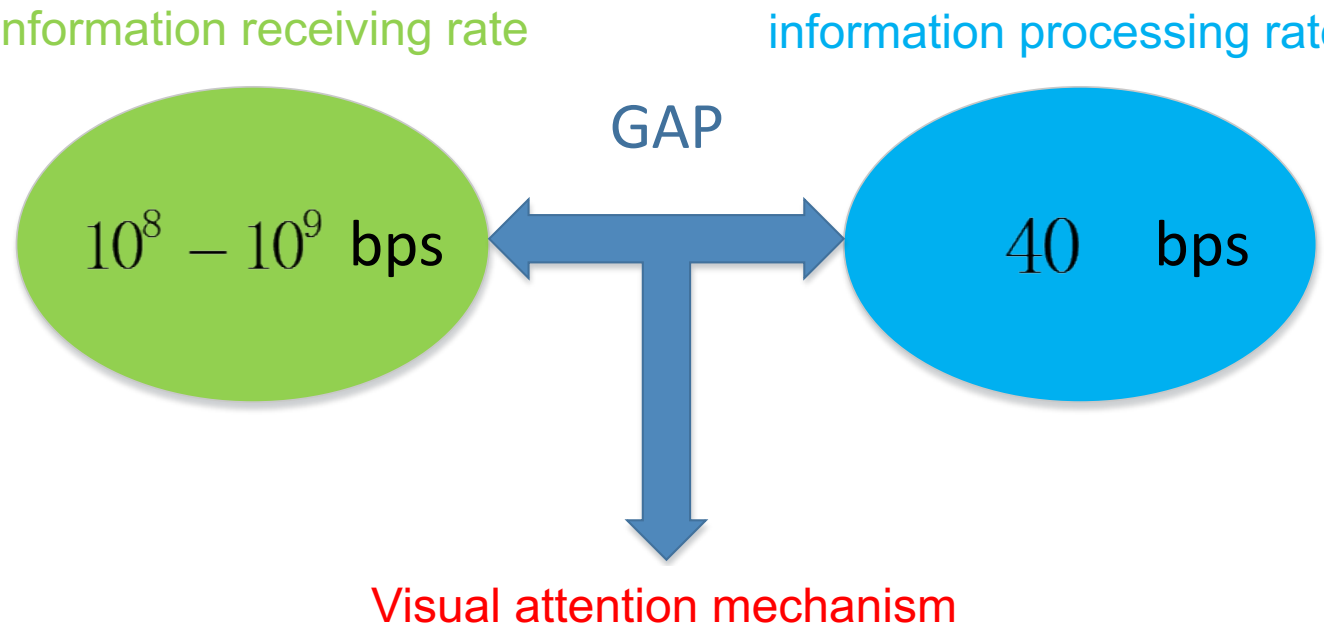




Where is Waldo, the young boy wearing the red-striped shirt...

Introduction to visual attention

- HVS
  - Amount of information coming down the optic nerve $10^8 - 10^9$ bits per second



Introduction to visual attention

- Visual attention
  - Posner proposed the following definition (Posner, 1980). Visual attention is used:
    - to select important areas of our visual field (alerting);
    - to search for a target in cluttered scenes (searching).
  - There are several kinds of visual attention:
    - Overt visual attention: involving eye movements;
    - Covert visual attention: without eye movements (Covert fixations are not observable).

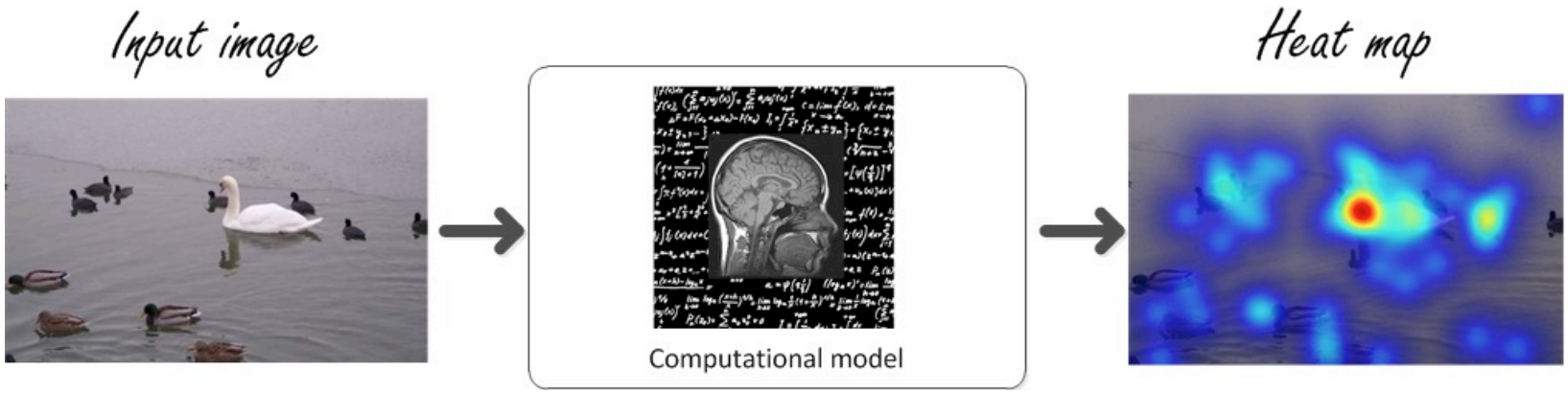
Introduction to visual attention

- Bottom-Up vs Top-Down
  - Bottom-Up: some things draw attention reflexively, in a task-independent way (Involuntary; Very quick; Unconscious);

  - Top-Down: some things draw volitional attention, in a task-dependent way (Voluntary; Very slow; Conscious).

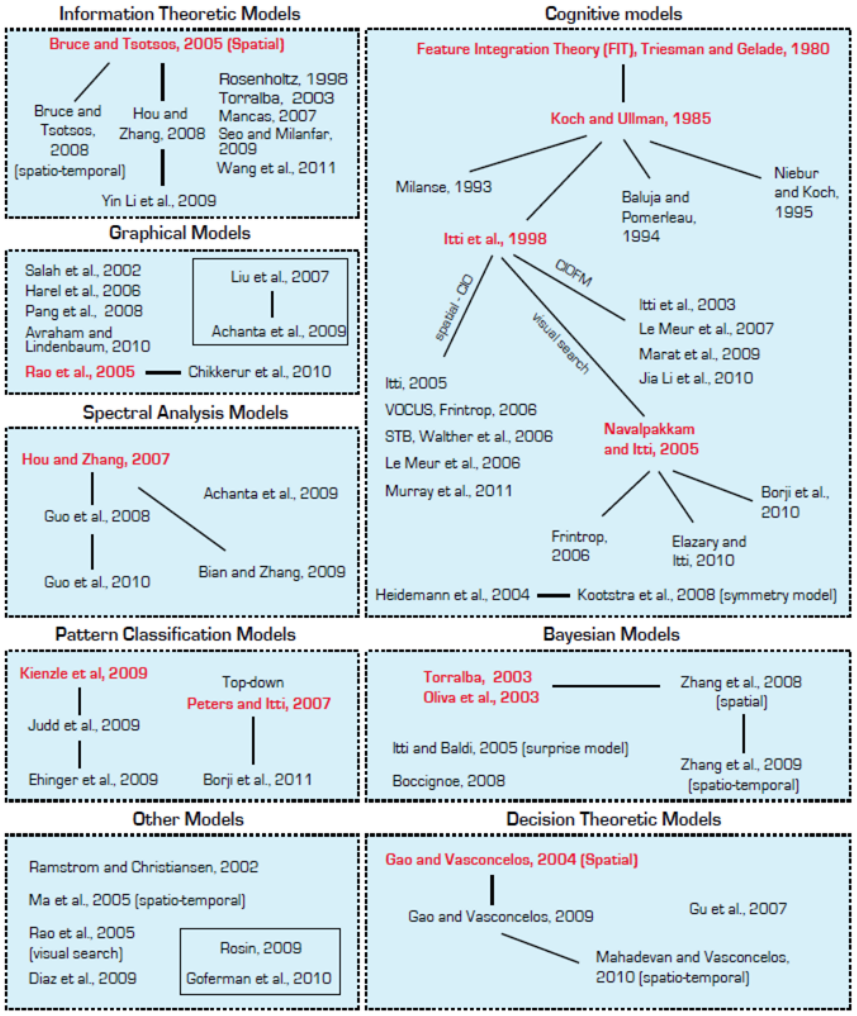
Introduction to visual attention

- Computational models of visual attention aim at predicting where we look within a scene.

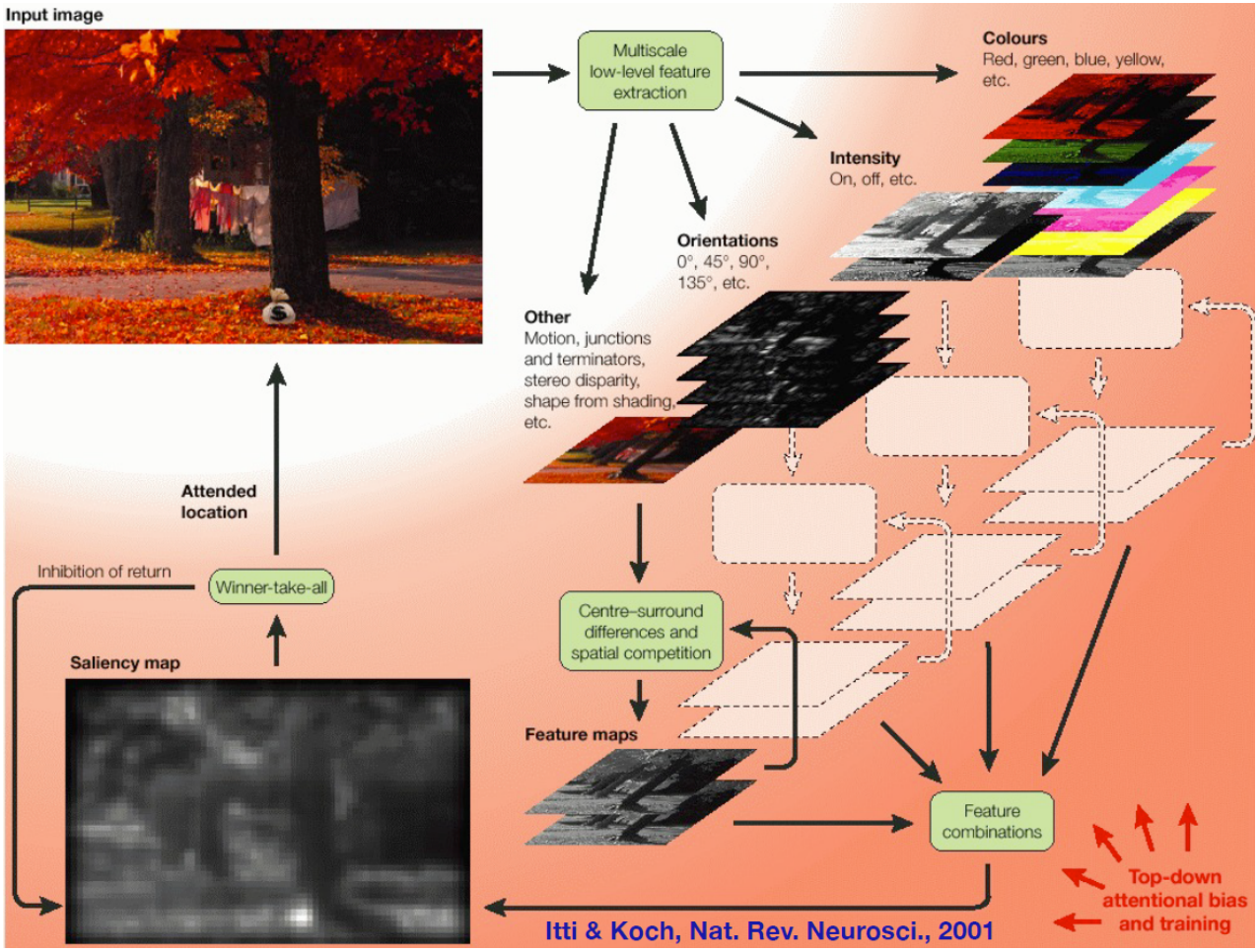


Computational models of Bottom-up visual attention

- Taxonomy of models:
  - Information Theoretic models;
  - Cognitive models;
  - Graphical models;
  - Spectral analysis models;
  - Pattern classification models;
  - Bayesian models.
  - Deep network-based models.

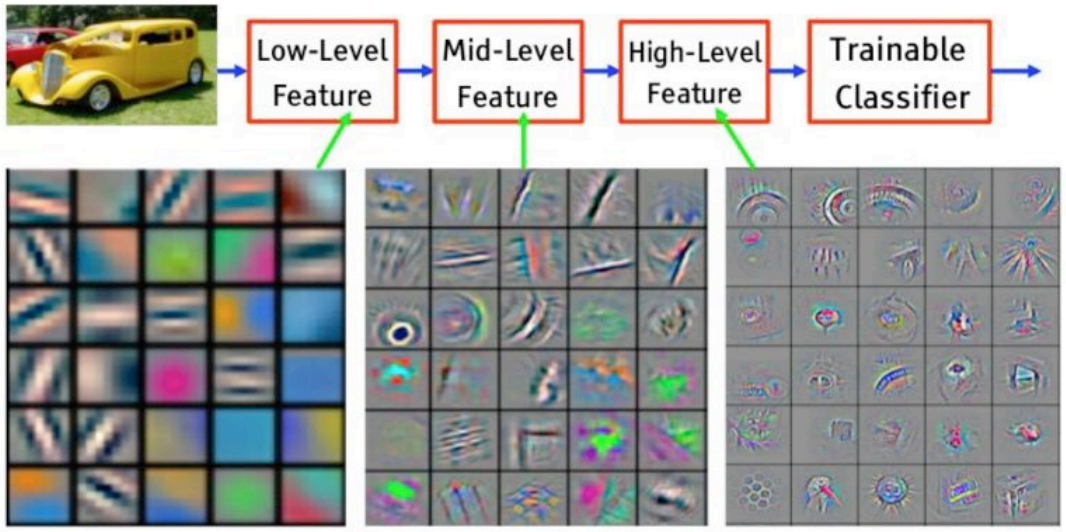


Computational models of Bottom-up visual attention



CNN

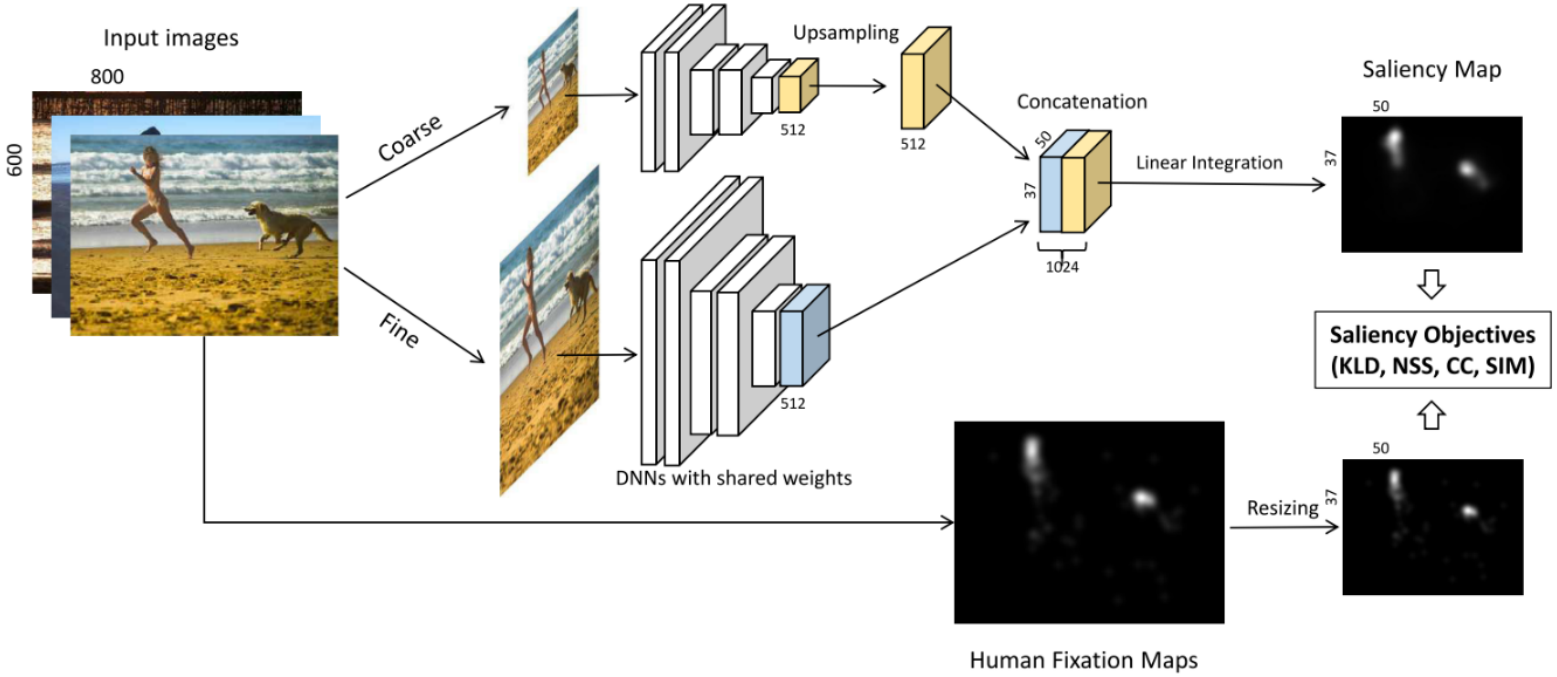
- A neural network model is a series of hierarchically connected functions;
- Each function’ s output is the input for the next function;
- These functions produce features of higher and higher abstractions;



- End-to-end learning of feature hierarchies.

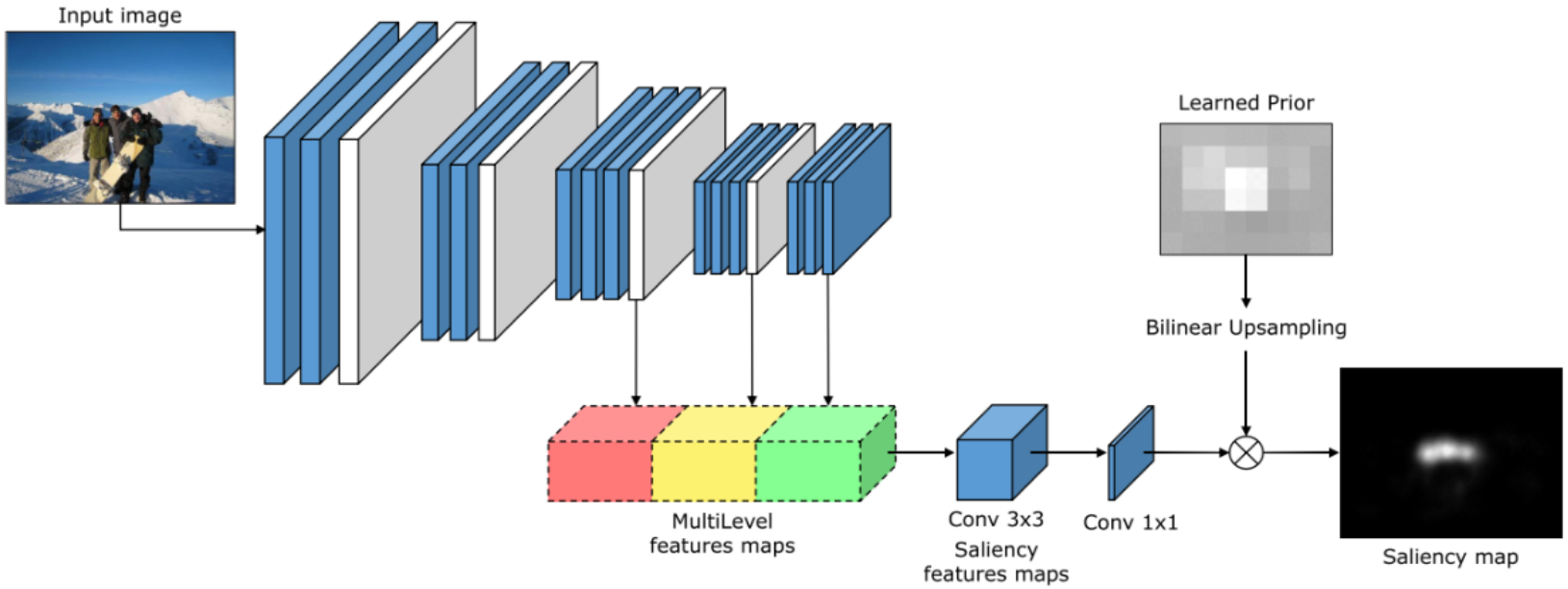
Image courtesy: http://www.iro.umontreal.ca/~bengioy/talks/DL-Tutorial-NIPS2015.pdf

Deep network: SALICON



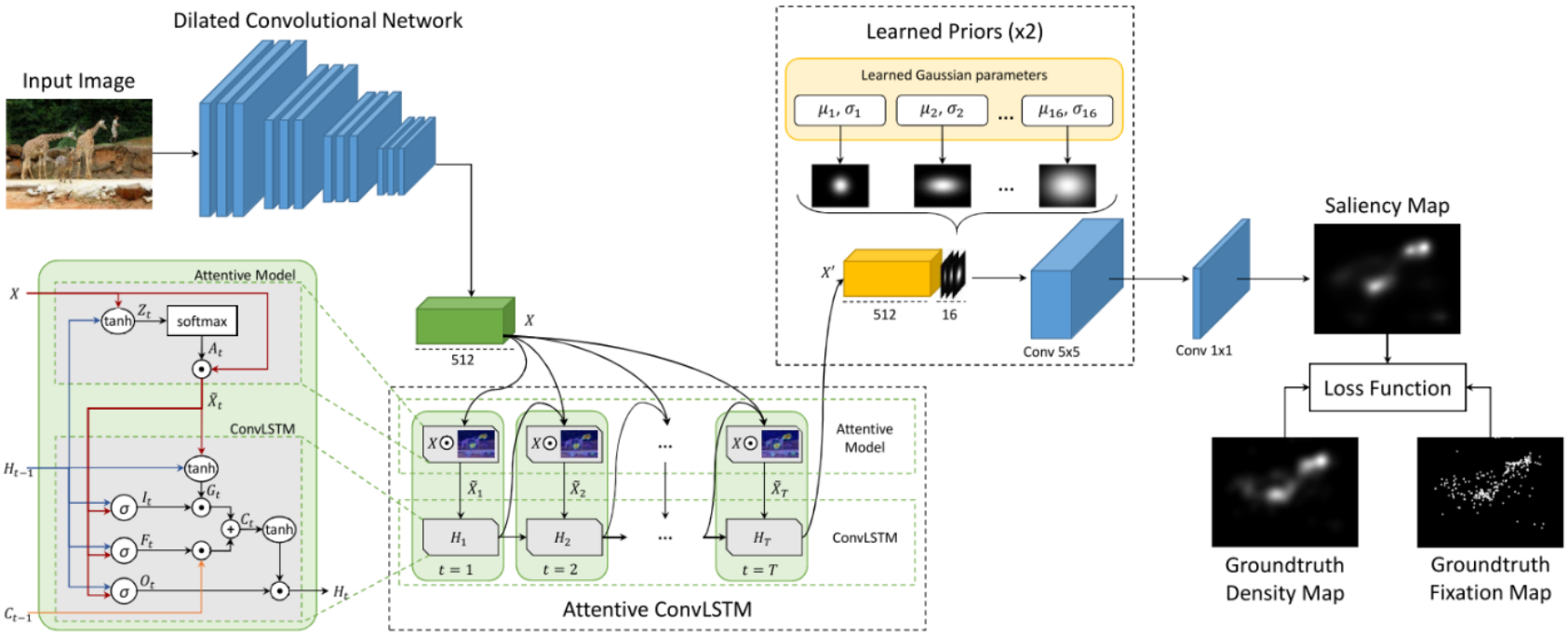
"Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks," ICCV 2015.

Deep network: ML-NET



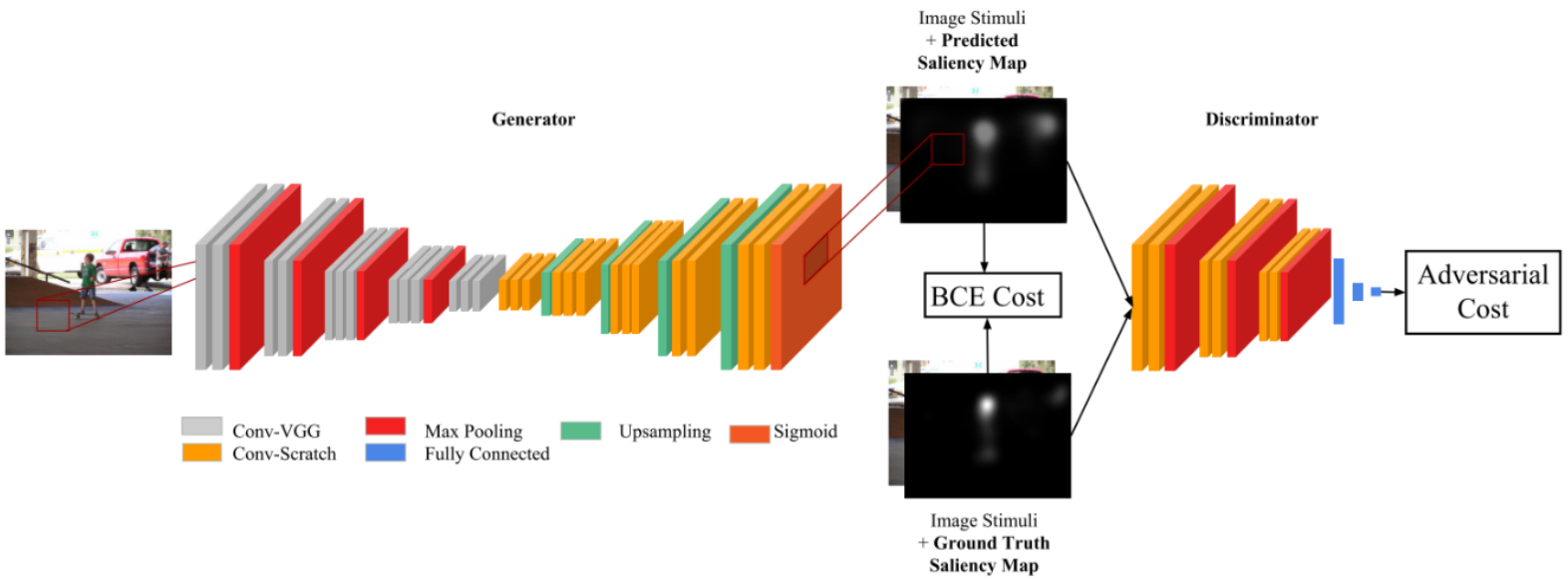
“A Deep Multi-Level Network for Saliency Prediction” ICPR 2016.

Deep network: SAM-VGG & SAM-ResNet



"Predicting human eye fixations via an lstm-based saliency attentive model" TIP 2018.

Deep network: SalGAN



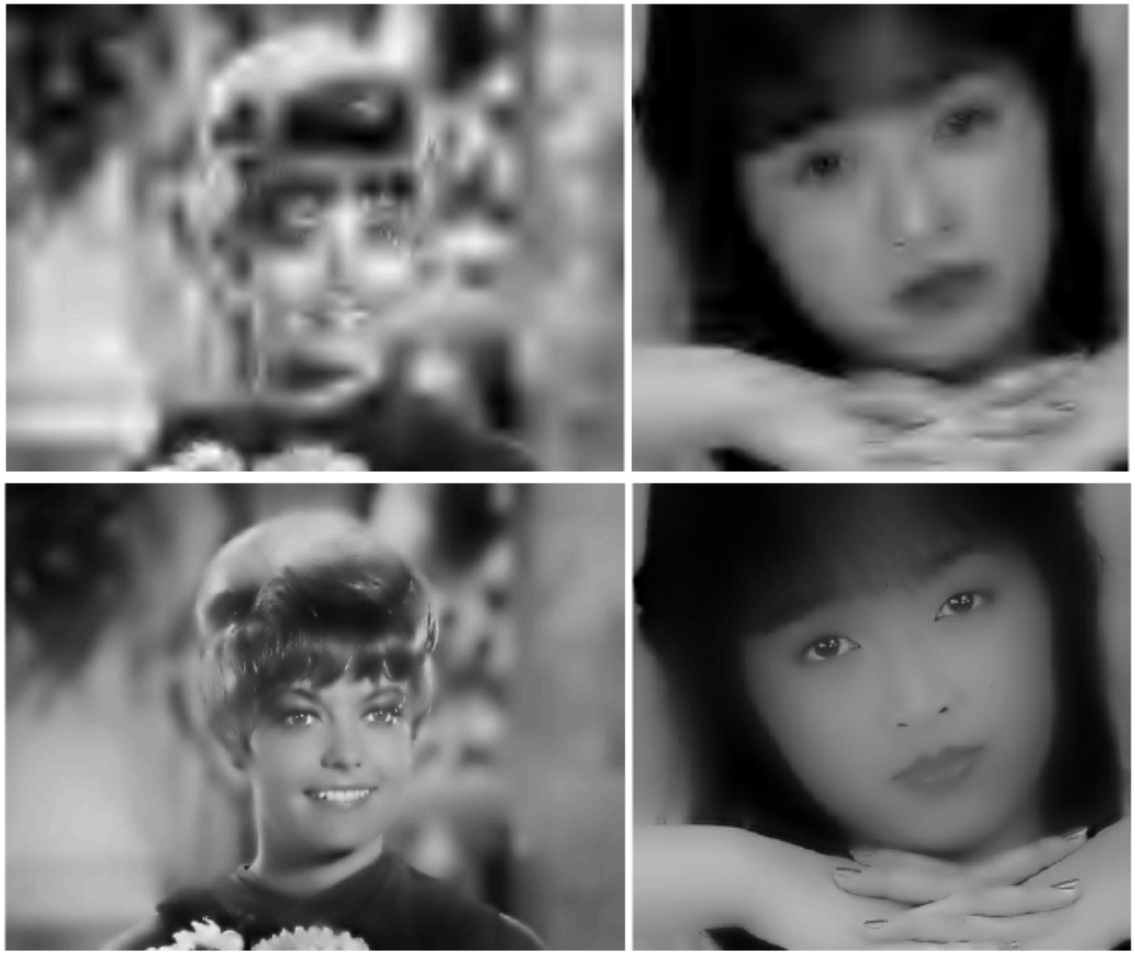
“Salgan: Visual saliency prediction with generative adversarial networks” CVPR 2017.

Saliency-based applications

- Image scaling based on image content

Saliency-based applications

- Compression and quality evaluation



Saliency-based applications

- Diagnosis of neurodevelopmental disorders (see Itti, L. (2015). *New Eye-Tracking Techniques May Revolutionize Mental Health Screening*. Neuron, 88(3), 442-444.);
- Learning Visual Attention to Identify People With Autism Spectrum Disorder (Jiang and Zhao, 2017);
- Alzheimer’ s disease (Crawford et al., 2015);
- US startup proposes a device for tracking your eyes to see if you’ re lying...;
- Emotion, gender (Coutrot et al., 2016), age (Le Meur et al., 2017)....

Contents

1 Background

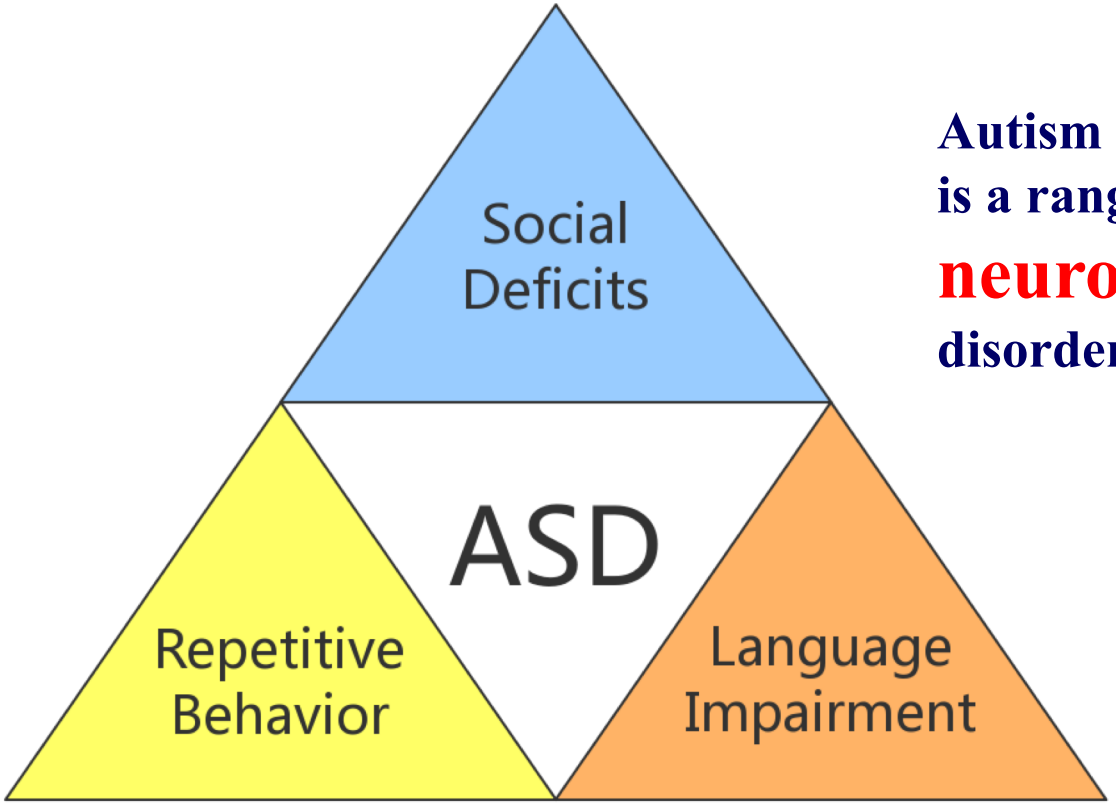
2 ICME2019 Grand Challenge

3 Solutions and Results

4 Assignment Requirements

BACKGROUND: ASD

BACKGROUND: ASD



Autism spectrum disorder (ASD) is a range of conditions classified as neurodevelopmental disorders, not psychological problem.

BACKGROUND: ASD

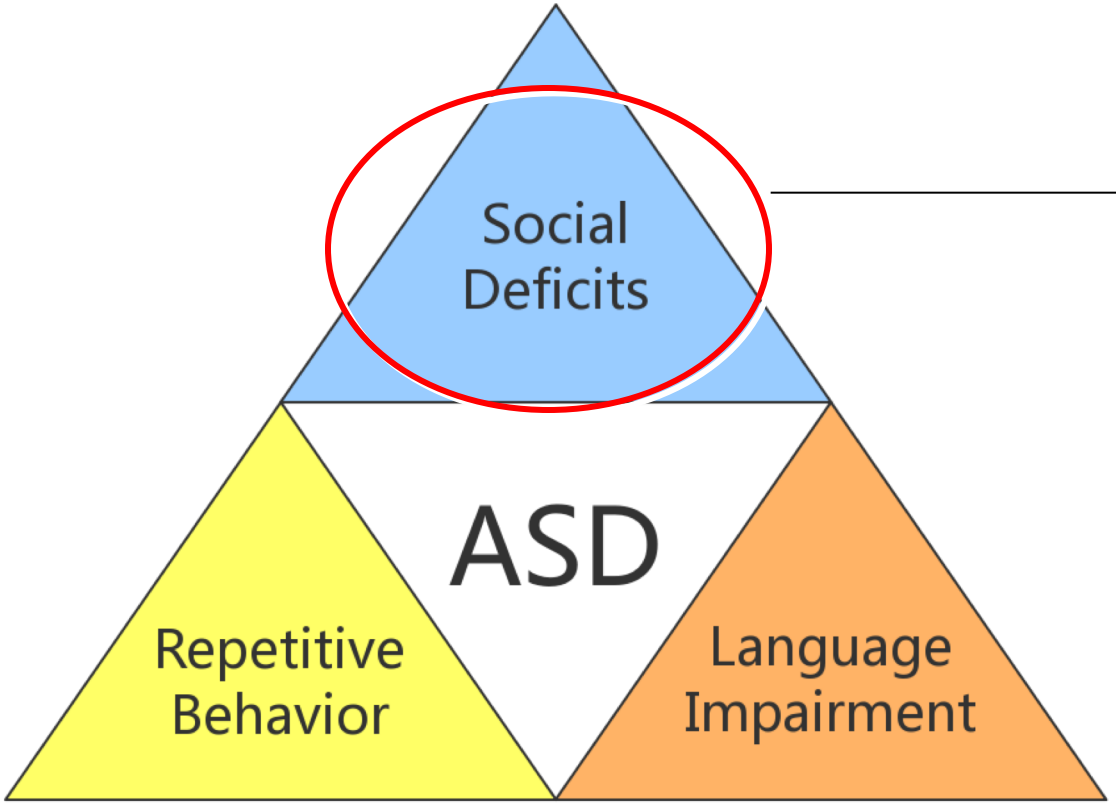
Identified Prevalence of Autism Spectrum Disorder

ADDM Network 2000-2014 Combining Data from All Sites

Surveillance Year	Birth Year	Number of ADDM Sites Reporting	Prevalence per 1,000 Children (Range)	This is about 1 in X children…
2000	1992	6	6.7 (4.5-9.9)	1 in 150
2002	1994	14	6.6 (3.3-10.6)	1 in 150
2004	1996	8	8.0 (4.6-9.8)	1 in 125
2006	1998	11	9.0 (4.2-12.1)	1 in 110
2008	2000	14	11.3 (4.8-21.2)	1 in 88
2010	2002	11	14.7 (5.7-21.9)	1 in 68
2012	2004	11	14.6 (8.2-24.6)	1 in 68
2014	2006	11	16.8 (13.1-29.3)	1 in 59

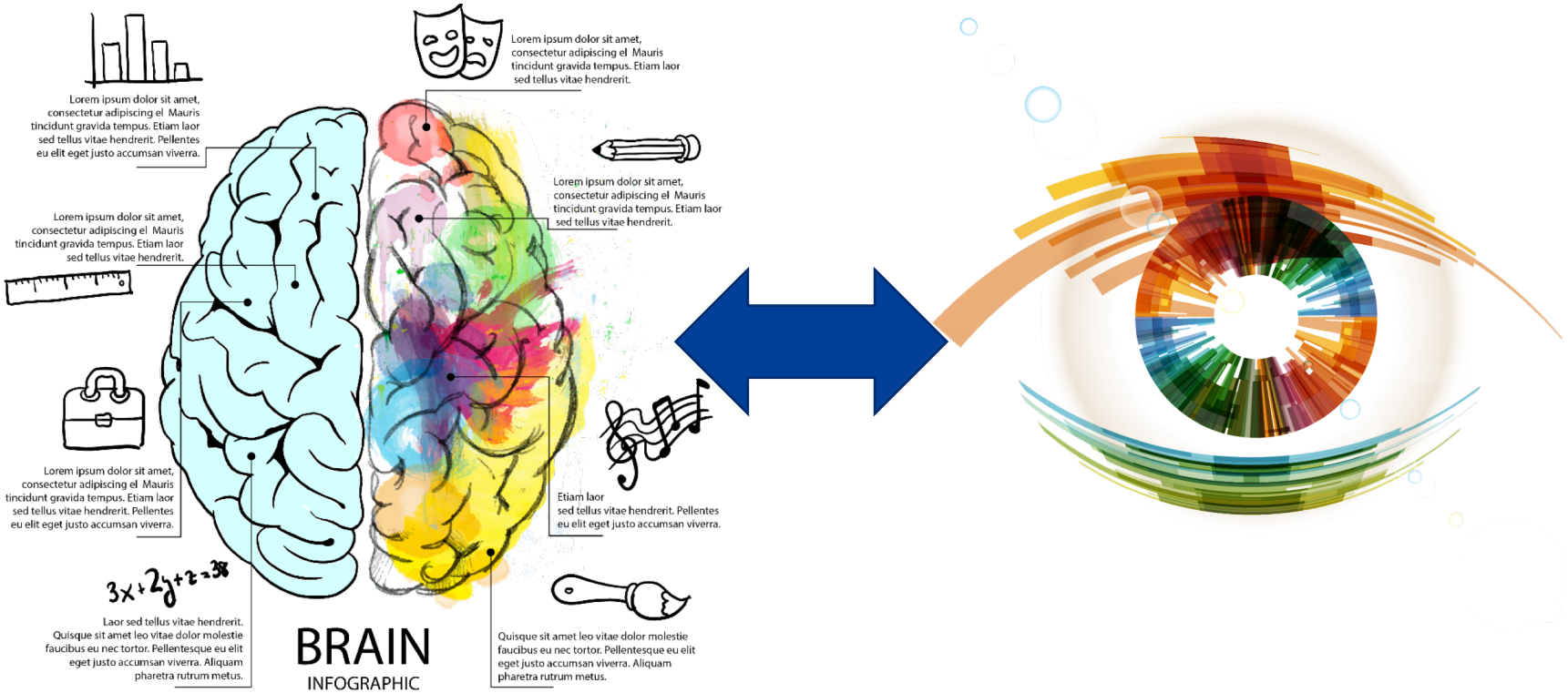
Reference: Centers for Disease Control and Prevention (https://www.cdc.gov/ncbddd/autism/data.html)

BACKGROUND: Why attention mechanism?



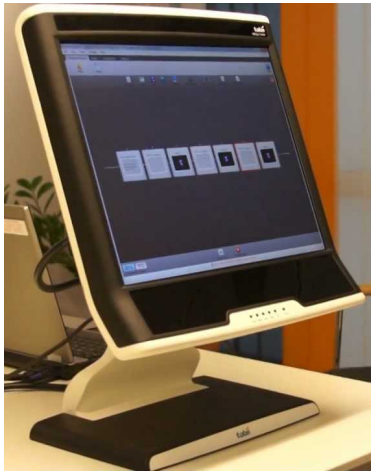
visual attention abnormal

BACKGROUND: Why attention mechanism?



Ground truth

- The requirement of a ground truth
  - Eye tracker (sampling frequency, accuracy…);
  - A panel of observers (age, naïve vs expert, men vs women…);
  - An appropriate protocol (free-viewing, task…).



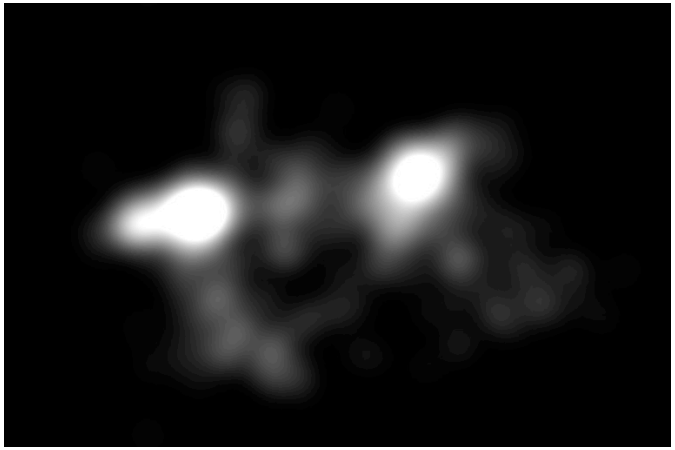
Ground truth

- Discrete fixation map:

$$f^i(\mathbf{x}) = \sum_{k=1}^{M} \delta(\mathbf{x} - \mathbf{x}_k)$$

- Continuous saliency map

$$S(\mathbf{x}) = \left(\frac{1}{N} \sum_{i=1}^{N} f^i(\mathbf{x}) \right) * G_\sigma(\mathbf{x})$$



ICME GC

USE AND ACCESS

- **Develop computational models** that:
  - **Fit gaze behavior** of people with ASD (saliency prediction or saccadic models).
  - Are able to **classify ASD/TD viewers** using gaze data.
- Dataset used in the **Grand Challenge *Saliency4ASD***: https://saliency4asd.ls2n.fr -- Email: saliency4ASD@ls2n.fr

SUBJECTIVE EXPERIMENT

EQUIPMENT: Tobii T120 Eye Tracker

- 17" display (1280x1024), sampling rate: 120 Hz.
- Viewing distance: 65 cm (Tracking range: 50-80 cm.)

SUBJECTS:

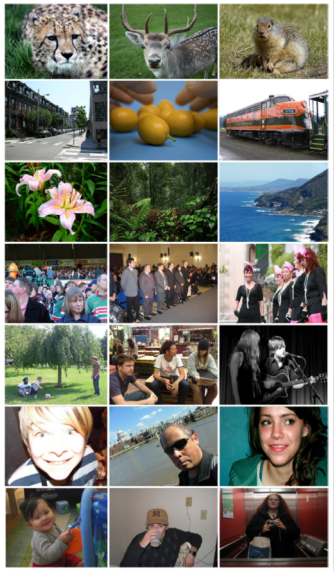
- 14 children with ASD and 14 healthy children (controls).
- Ages ranged from 5 to 12 years (avg.: 8 years).
- All with normal or correct-to-normal visual acuity.
- Matched gender, race, education, etc. of the two groups

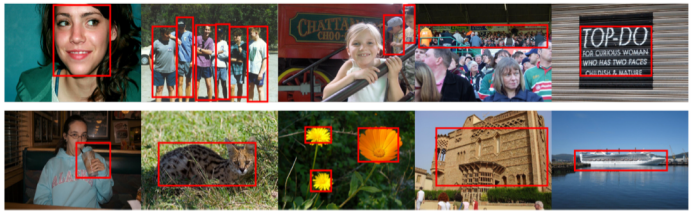
PROCEDURE: Look freely to the images

- Images showed in random order.
- Test split into 10 sessions (30 images each).
- Images showed at their full resolution for 3 seconds.
- 1 second grey-screen between images.
- Eye-tracking calibration before each session.

STIMULI

- **300 images** from MIT dataset [T. Judd *et al.* "Learning to predict where humans look", ICCV2009].
- **Content variety**: 7 categories
  - Animals (40), buildings or objects (88), natural scenes (20), multiple people (36), multiple people + objects (41), single person (32), single person + objects (43).
- **Annotated semantic-level features**: 10 categories
  - Faces (22), people (245), background people (14), crowd (34), texts (27), handheld objects (55), animals (70), plants (29), buildings (53), and objects (173).

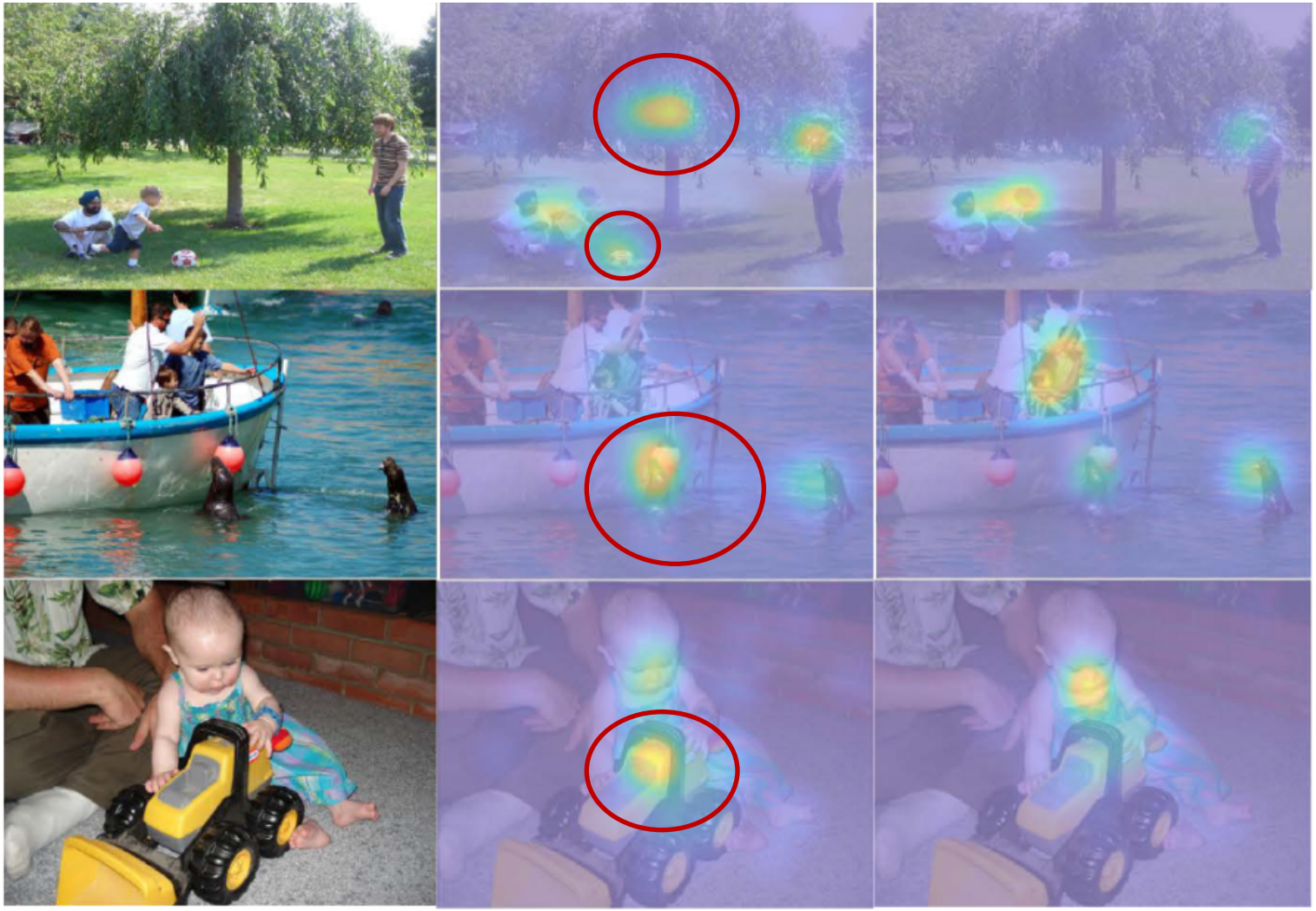




Difference Visualization

ASD

Healthy



Difference Visualization

Absence of joint attention

Lower social gaze

Hand bias

Anti-center bias

Evaluation metrics

mit saliency benchmark

about | results | datasets | submission | downloads

mit300 | cat2000

mit saliency benchmark results: mit300

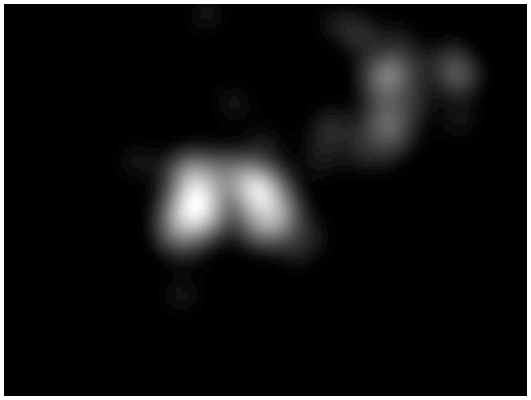
The following are results of models evaluated on their ability to predict ground truth human fixations on our **benchmark data set containing 300 natural images with eye tracking data from 39 observers**. We post the results here and provide a way for people to submit new models for evaluation.

Model Name	Published	Code	AUC-Judd [?]	SIM [?]	EMD [?]	AUC-Borji [?]	sAUC [?]	CC [?]	NSS [?]	KL [?]	Date tested [key]	Sample [img]
Baseline: infinite humans [?]			0.92	1	0	0.88	0.81	1	3.29	0		

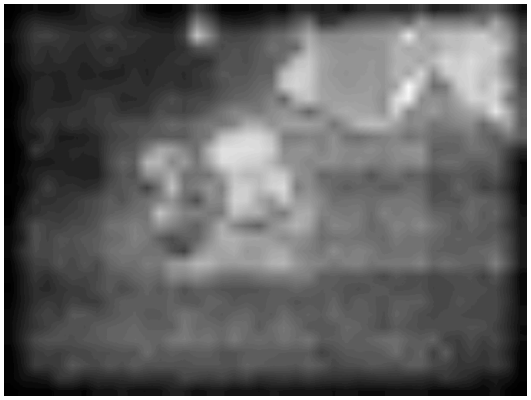
AUC-Judd, SIM, EMD, AUC-Borji, sAUC, CC, NSS, KL

Evaluation metrics—— AUC

(a) Original



(b) Human



(c) Itti’ s model

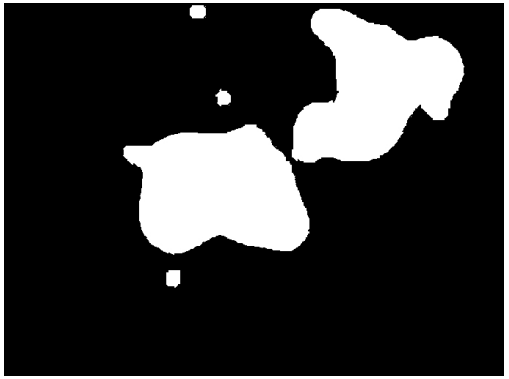
(1) Label the pixels of the human map as fixated (255) or not (0):

The threshold is often arbitrary chosen (to cover around 20% of the picture).

Evaluation metrics—— AUC

(2) Label the pixels of the predicted map as fixated (255) or not (0) by a given threshold

(3) Count the good and bad predictions between human and predicted maps:

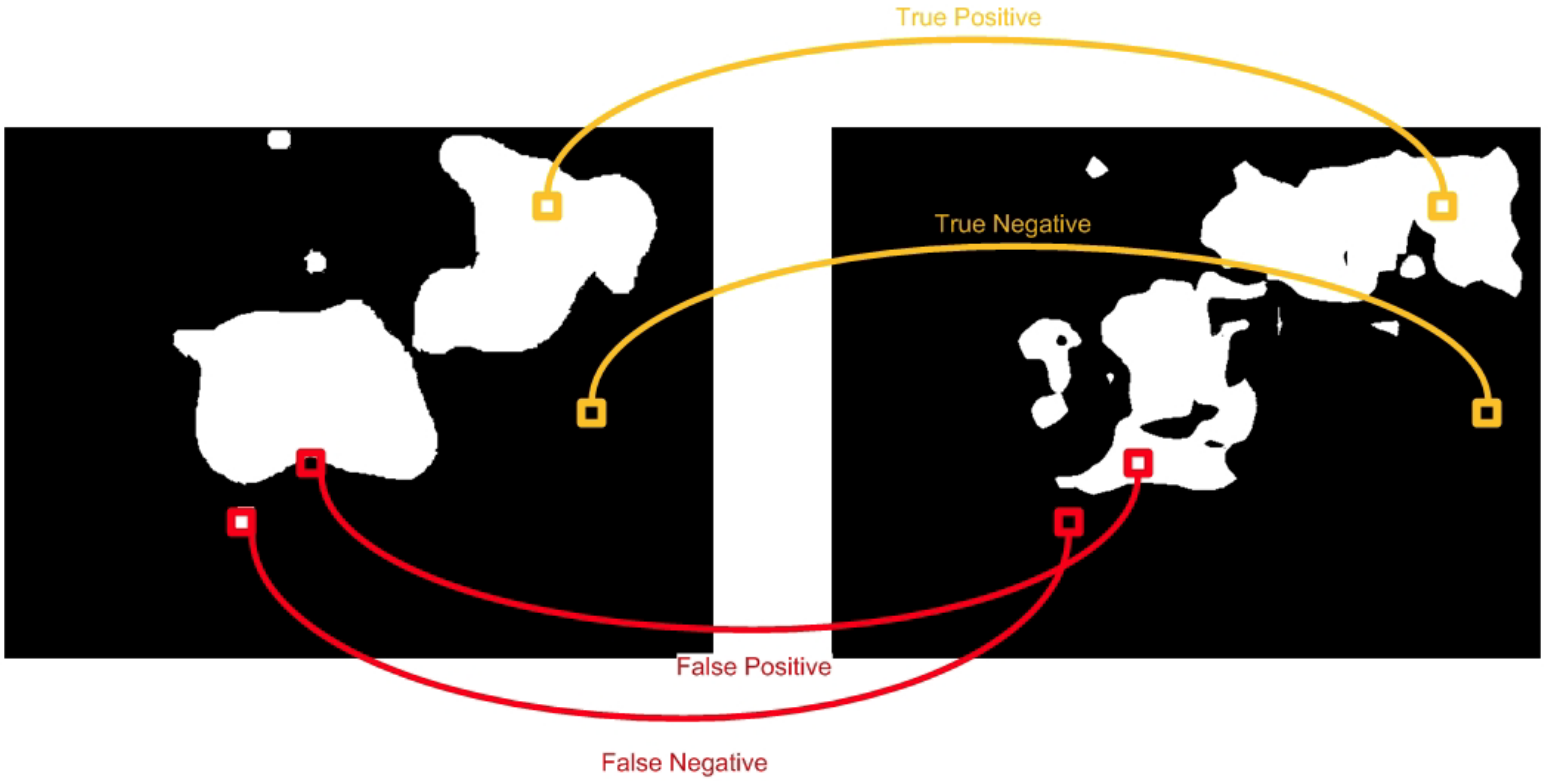


(a) Human Bin.

(b) Predicted Bin.

Evaluation metrics—— AUC

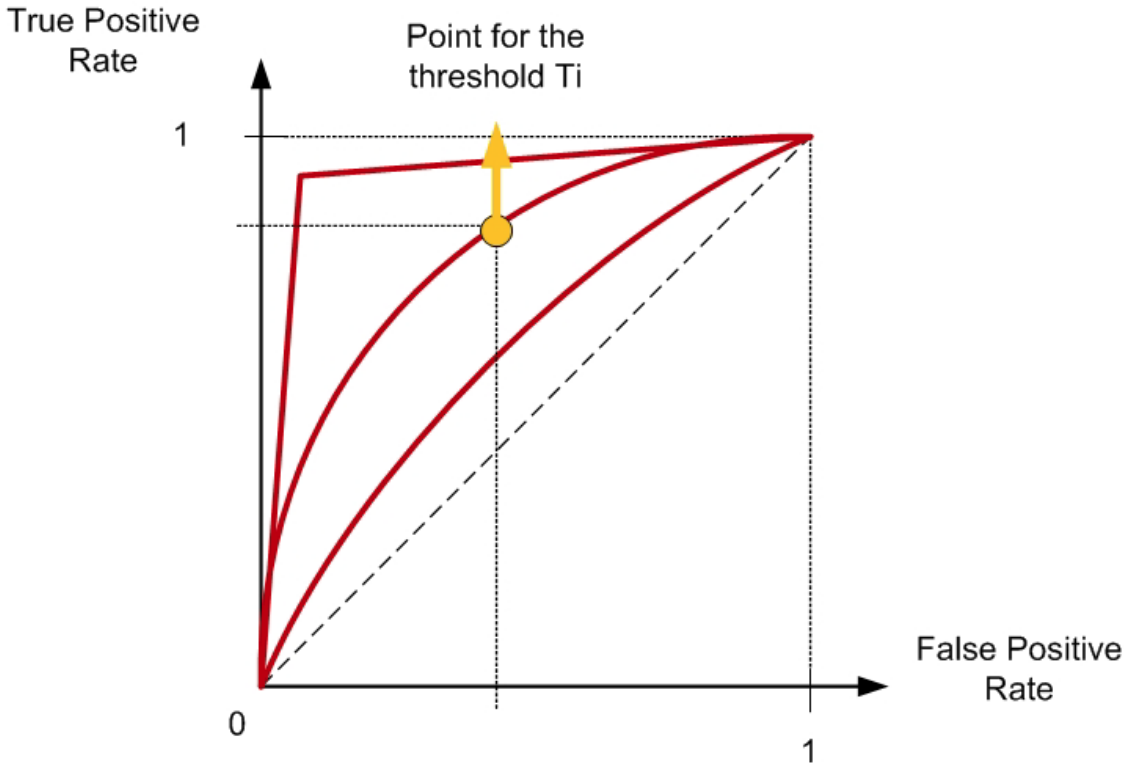
(3) Count the good and bad predictions between human and predicted maps:



True Positive Rate = True Positive / (True Positive+False Negative)
False Positive Rate = False Positive / (False Positive+True Negative)

Evaluation metrics—— AUC

(4) Go back to (2) to use another threshold… Stop the process when all thresholds are tested.



AUC (Area Under Curve)

Contents

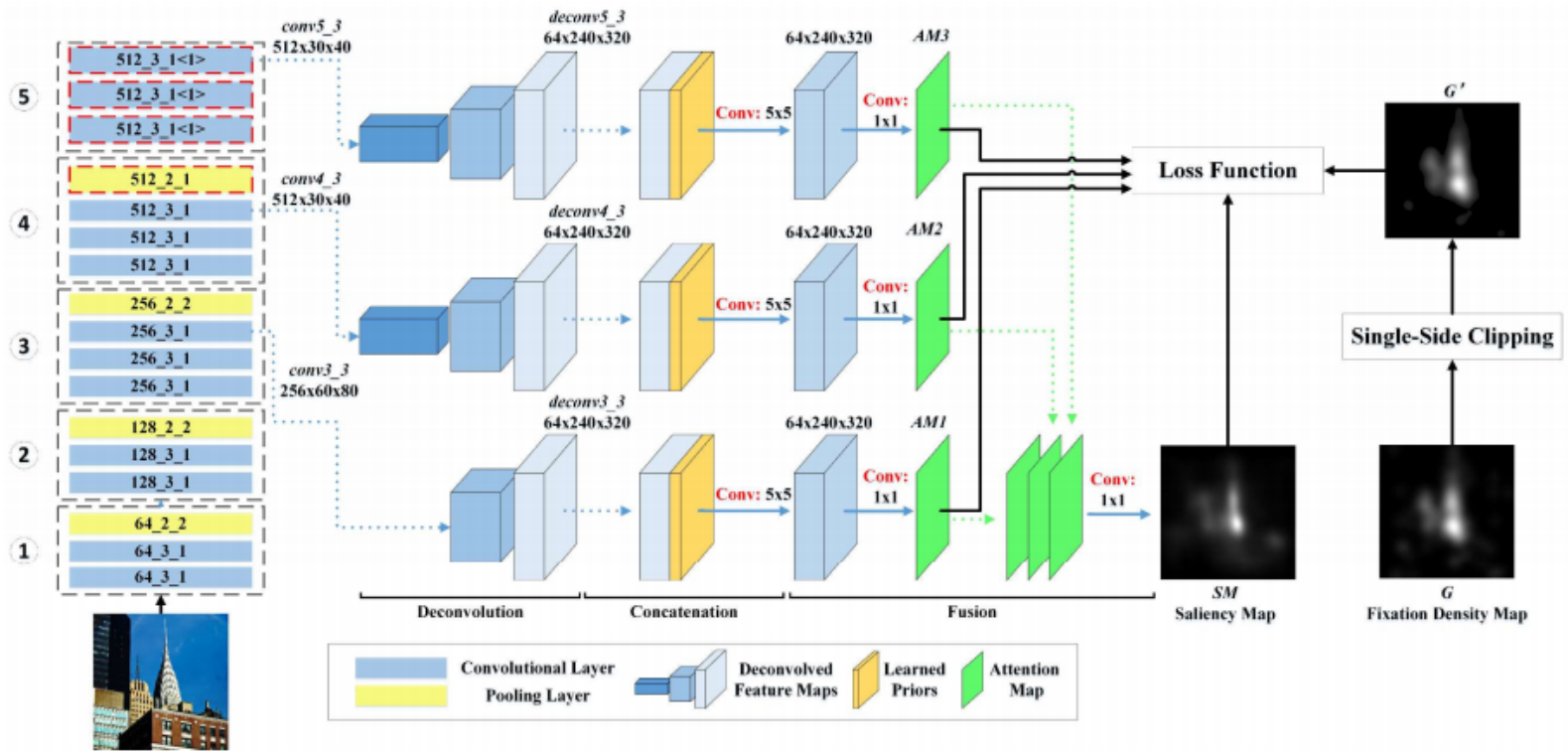
1 Background

2 ICME2019 Grand Challenge

3 Solutions and Results

4 Assignment Requirements

Saliency Prediction



Classifying

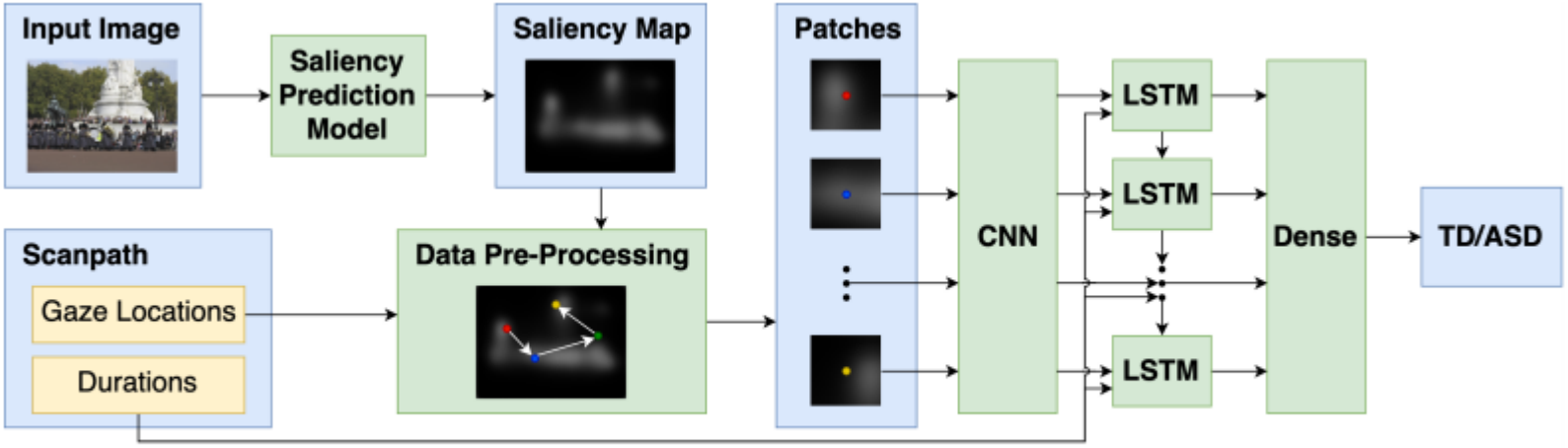


Fig. 1. The proposed SP-ASDNet framework.

Contents

1 Background

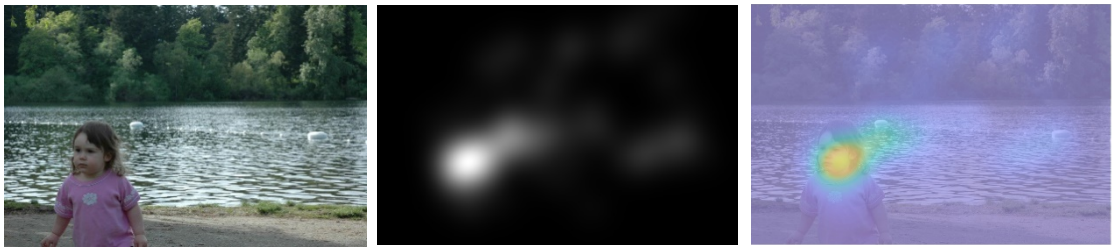
2 ICME2019 Grand Challenge

3 Solutions and Results

4 Assignment Requirements

Assignment Requirements

1. Design a **saliency prediction model** that fits gaze behavior of people with ASD, by using the datasets released by the ICME Grand Challenge.
2. **Write a report** in IEEE Journal format. The report should include title, abstract, the implementation details of your models, the experimental setup and the experimental results, and some analyses, etc.
3. Validate the proposed model on the test dataset (report AUC, sAUC, CC, NSS).
4. (Additional task: if you are interested in the second track, you can also try to classify the ASD individuals and TD people from gaze data.)



Finally, you should send the report, the model (including the readme file), and the results on the test dataset to sunguwei@sjtu.edu.cn. The email title should be written as: Assignment1 + Name1 + StudentNumber1 + Name2 + StudentNumber2 ……

Thank you !