Modeling human attention by learning from large amount of emotional images

Macario O. Cordel II
Advanced Research Institute for Informatics, Computing and Networking
Data Science Institute, De La Salle University
Manila, Philippines
macario.cordel@dlsu.edu.ph

***Abstract*—Recent resurgence of neural networks in computer vision have resulted in tremendous improvements in saliency prediction, eventually, saturating some saliency metrics. This leads researchers to devise higher-level concepts in images in order to match the key image regions attended to by human observers. In this paper, we propose a saliency model which utilizes the top-down attention mechanism through the involvement of emotion-inducing region information in the predictor's feature space. The proposed framework is inspired by psychological and neurological studies that emotion attracts attention. Using three publicly available datasets with emotion-rich images, we were able to show that awareness of the emotion-inducing region improves saliency prediction of images. Saliency metrics for probabilistic models, particularly information gain and KL-divergence, have improved with respect to the same architecture without emotion information. Statistical tests show that emotional regions generally have higher improvement than neutral regions corroborating psychological studies that emotion attracts attention.**

***Index Terms*—visual saliency model, emotion stimuli map, attention, deep learning**

I. Introduction

Predicting where human eyes will fixate has attracted a significant number of researchers because of its potential applications such as in human-computer interaction and robot vision. Traditionally, image saliency prediction algorithms use hand-designed features which make the feature space very limited. The recent success of deep neural networks (DNN) in saliency prediction models, have demonstrated that such task requires intricate feature space. In fact, most of the best performing algorithms in [1] are DNN-based, with marginal performance over other similar DNN-based systems, saturating some evaluation metrics. Recently, Bylinskii et al. [2] re-examined current saliency prediction algorithms, and argued that to approach human-level performance, saliency models will need to discover higher-level concepts in images, such as text or motion, and reason about the relative importance of image regions. One possible high level concept in images is emotion, a complex but well-studied determinant in human attention [3]–[6]. Behavioral observations show that people pay attention to affective rather than neutral stimuli, and this commonly happens spontaneously [4]. In fact, in a visual search task, the objects can be easily found if it contains affective values [5], [6], e.g. a bloody knife in a bedroom or a snake among flowers. Conversely, Ren et al [7] showed that emotion can be extracted from visual stimuli.

978-1-7281-0858-2/19/$31.00 ©2019 IEEE

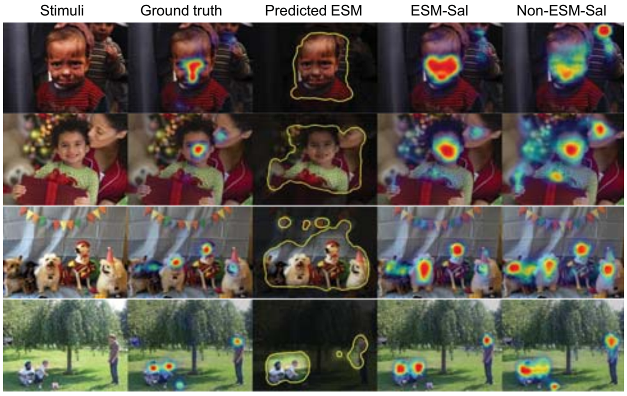

Fig. 1. Studies show that people pay attention to emotional objects rather than neutral objects. We show the improvement in saliency prediction when the feature detectors are trained to be aware of the affective region in an image. For illustration purposes, the affective region (third column) is marked with yellow line. The Non-ESM-Sal column shows the saliency prediction trained with no emotion while the ESM-Sal column shows the saliency prediction with emotion. In the examples presented, the emotion-attentive model tend to show improvement in relative saliency of objects and mis-detections.

Motivated by these, we investigate if emotion-inducing region in an image would improve the performance of visual saliency prediction. It is therefore our objective (1) to introduce the emotion-inducing region being the emotion information, as a potential higher-level image concept, to the saliency model's feature space and (2) to perform fine-grained evaluation using appropriate saliency metrics and saliency dataset containing emotional images, visualization of relative saliency of emotional and neutral image regions. We hypothesize that the emotion-inducing region have implicit attraction to human attention akin to psychological observations; and exploiting this information would provide improvement in saliency prediction of images with highly affective objects. We were able to show that emotion information helps improve saliency prediction for image types with emotional focal objects as shown in Fig. 1.

In this paper, we show that a saliency model with emotionally developed feature detectors (i.e. Emotion Stimuli Map-aware Saliency model or ESM-Sal) has less misdetection and better relative saliency prediction as provided by the improved IG and KL saliency metrics, compared to saliency model trained only on low-level object semantic (Non-Emotion Stimuli Map-aware Saliency model or Non-ESM-Sal). Here, we introduce a scheme for introducing the emotion information to the saliency model via simultaneous training the feature detectors using fixation map and emotion stimuli map (ESM) [8]. ESM is the pixel-wise contribution to the evoked emotion representing the image's emotion-inducing region.

II. Related work

Mimicking the behaviour of how human gazes and instinctively focuses his attention have been the template for solving the saliency prediction problem. The traditional algorithms for saliency predictions [9]–[12], for example, are based on the Feature Integration Theory of attention [13] which suggests that features (e.g. color, orientation, spatial frequency etc.) are registered early, automatically and in parallel across the visual field, while objects are identified separately and only at a later stage. In recent DNN-based saliency models, the human's selective attention at different resolution, influences the architecture of SALICON [14] which uses two branches of feature detectors to allow fine and coarse resolution as input to their saliency model. DeepFix [15] leveraged on the selective attention, as well, but uses inception-style convolution blocks. Some state-of-the-art systems also incorporated the central fixation bias of humans in their system either by direct superimposing the center priors e.g. in DeepGaze I [16], and DeepGaze II [17], or by including the location bias in the learning phase e.g. SAM [18]. These high-performing saliency prediction models took advantage on the representational power of semantic-rich Deep Neural Network (DNN) feature detectors e.g. VGG-16 [19] and GoogleNet [20].

Emotion is one of the top-down cues which plays an important role in human perception [3] and its interaction with attention is well-established in different neuroscience and psychological studies [21]–[24]. In computer science, Peng et al [8] and Sun et al [25] both introduced systems which predict the affective region (AR) in an image. Peng et al proposed the prediction of Emotion Stimuli Map (ESM) problem which estimates the pixel-wise contribution to evoked emotion of an image. These two works were able to show that emotion-inducing region could be actually predicted. However, the resulting emotion-inducing map certainly does not reflect the visual saliency of an image.

In addition to these works, the interaction between the visual saliency and visual sentiment was also investigated in the works of Zheng et al [26]. Rather than generating a saliency map, the authors used the latter to determine possible emotional objects in the image and perform sentiment classification. Each object are then compared to the overall image sentiment to measure the object emotion and image sentiment agreement. The image is further categorized into e.g. indoor/outdoor, natural/man-made and face/no face, to perform analysis. Results show that image sentiments are mainly influenced by the outstanding presence of man-made objects or faces or are in indoor scenes. In this work, we will show that emotion-inducing regions in images are those with better saliency prediction.

III. Architecture

We present the Emotion Stimuli Map-based saliency model (ESM-Sal) which is illustrated in Figure 2. Our saliency model is a straight forward model composed of the VGG convolution layers as feature detector, followed by a cascade of two inception-style convolution blocks to generate feature maps of different resolution. The feature detector is shared with the Emotion Stimuli Map (ESM) prediction model such that the feature detector development is influenced by the emotion-inducing region in the stimuli. Sharing of feature detector is achieved by tapping the second and fourth convolution block layers of the feature detector through 256-feature map inception blocks and the fifth convolution block layer through a 512-feature map inception block. Note that we removed the max-pooling from the inception blocks as preliminary results show better performance if pooling is removed.

For the ESM prediction stream, all outputs from the inception blocks are down sampled to 25×19 size to match the fifth layer feature maps. A 1024-3x3 convolution layer is used to increase the non-linearity in ESM prediction, before the 1024 feature maps are reduced to a 2D, 25×19 output ESM. For the saliency prediction stream, a layer of 1×1 convolution layer is used after the inception layer to down-sample the 512 feature maps to a single 25×19 output saliency map.

IV. Experiments

The ESM-Sal model is designed for images containing emotion-inducing regions. We evaluated our proposed saliency model on three datasets rich in affective images. The first dataset is the NUSEF [28] which is a public eye fixation database composed of 758 images, 287 of which came from the International Affective Picture System (IAPS) [29]. Out of all the 758 images in NUSEF, 383 of these carry emotion while the rest are concepts such as indoor-outdoor scenes, and living and non-living things. The second dataset is the EMOtional attention dataset (EMOd) [30] which consists of 1249 images, 389 of which are also from IAPS. Finally, the third dataset is the CAT2000 [31] which is composed of 2000 images from 20 different categories, varying from natural indoor to outdoor scenes to artificial stimuli like cartoons and line drawing [31]. We used the Affective category of CAT2000 which contains emotional images.

A. Training and testing

Our proposed saliency prediction model is trained simultaneously with the ESM prediction model, to allow the feature detectors to learn the semantic information that locates affective region. The pre-trained VGG-16 is used as the main feature detector and then the whole saliency model is fine-tuned using the SALICON training dataset with a momentum of 0.9 and initial learning rate equal to 1e-5. The learning rate decreases every 3e+4 iterations at a rate of 0.1. Due to the

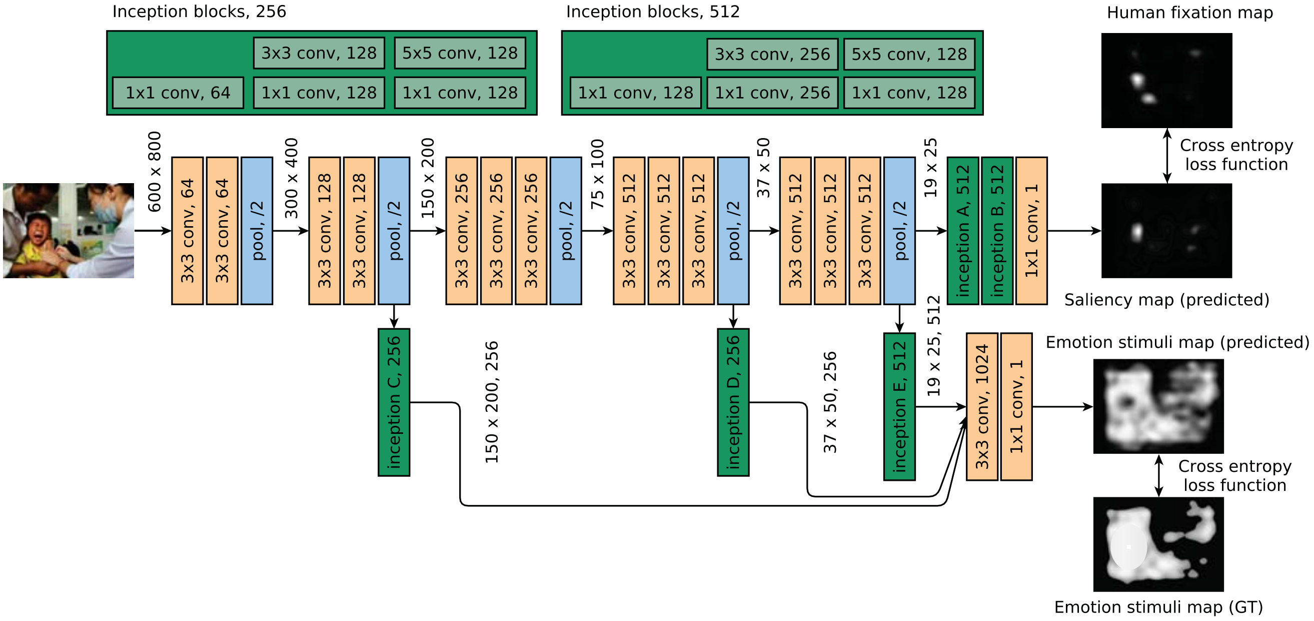

Fig. 2. A key challenge in introducing emotion in saliency prediction is integrating emotion information in the saliency model. Our proposed Emotion Stimuli Map-based saliency model does this by designing the saliency prediction to share its feature detectors with ESM prediction. The ESM prediction taps from the different layers (low, middle and high level) of the featured detector.

large of amount of training data and limited memory resources, training is performed one image per iteration. Validation data shows that after 3 epochs, the performance started to stabilize, but the training is performed for 6 epochs.

The ESM prediction branch identifies the affective region in the input image. As this branch shares the main feature detector, the feature detector for the saliency prediction in all likelihood becomes aware of emotional region and factors this information in saliency prediction. All inception-style convolution blocks, including those in the main saliency prediction branch, were initialized randomly with Gaussian distribution and standard deviation equal to 1e-4. The training of the saliency prediction model requires ground truth of both the fixation map and ESM, the ESM ground truth for the SALICON dataset were produced using the state-of-the-art ESM prediction system [8]. The fine-tuning was performed using GeForce GTX TITAN X.

B. Results

The quantitative results of the evaluation obtained on EMOd, NUSEF and Affective category of CAT2000 are presented in Table I. The values in bold for each column correspond to the two most desirable performance values across all saliency models. Comparing the changes between ESM-Sal and Non-ESM-Sal performances when the emotion information is incorporated in the saliency feature detector, decent improvement in KL and IG scores and decline in both EMD and NSS performances are consistently seen for the three datasets. Particularly, KL and IG scores are increased by 0.07 to 0.18 and 0.05 to 0.22, respectively; while EMD and NSS performances are worsened by 0.07 to 0.19 and 0.03 to 0.13, respectively. The AUC-Judd, AUC-Borji, sAUC, SIM and CC shows almost negligible score movement when emotion information is introduced to the model.

With respect to the other state-of-the-art algorithms, considering the straight forward design of ESM-Sal as compared to the parallel VGG networks of SALICON and the generative-adversarial network approach of SalGAN, the ESM-Sal performance in terms of IG and KL scores, is on par with other saliency models based on deep network.

For example, the performance margin in NUSEF dataset, between ESM-Sal performance and the state-of-the-art SALICON are decreased from 0.23 to 0.05 for KL and from 0.34 to 0.12 for IG. For EMOd, the KL and IG differences from the best KL and IG scores have slightly improved, from 0.10 to 0.05 for KL and from 0.16 to 0.11 for IG. Finally, for Affective CAT2000 images, KL and IG score differences from SALICON scores are decreased from 0.12 to 0.05 for KL and from 0.23 to 0.11 for IG. These improvement in KL and IG scores, deteriorates however the EMD and NSS scores.

The qualitative results of the evaluation obtained by our proposed work along with other saliency models on sample emotional images are shown in Figure 3. We added the last column Diff_Im which shows the difference between the ESM-Sal prediction and the Non-ESM-Sal prediction. The redness and the blueness of the pixels in the Diff_Im column shows the relative increase and decrease of saliency values of ESM-Sal with respect to Non-ESM-Sal.

The first three examples illustrate the qualitative improvement on the relative saliency prediction of ESM-Sal over

TABLE I

QUANTITATIVE COMPARISON OF SALIENCY SCORES OF THE EMOTION-ATTENTIVE SALIENCY PREDICTION AND OTHER SALIENCY MODEL (ESM-SAL) USING NUSEF, EMOD AND CAT2000 DATASET. THE HIGHLIGHTED VALUES ARE THE THREE MOST DESIRABLE VALUES IN EACH METRIC.

	Metrics	AUC-Judd ↑	AUC-Borji ↑	sAUC ↑	CC ↑	SIM ↑	EMD ↓	NSS ↑	KL ↓	IG ↑
NUSEF	ESM-Sal	0.81	**0.79**	0.75	0.63	**0.60**	2.76	1.45	0.64	0.62
	Non-ESM-Sal	0.81	0.76	0.74	0.63	0.59	2.62	1.44	0.82	0.40
	SALICON [14]	0.81	**0.79**	**0.76**	**0.66**	**0.60**	**2.45**	1.48	**0.59**	**0.74**
	SalGAN [27]	**0.83**	0.78	0.75	**0.66**	0.58	2.72	**1.72**	0.90	0.51
	SROD [12]	0.75	0.74	0.69	0.40	0.47	4.44	0.96	0.97	0.18
	BMS [10]	0.74	0.74	0.64	0.38	0.45	4.14	0.96	2.40	-0.17
	GBVS [11]	0.81	**0.79**	0.74	0.56	0.54	3.21	1.32	0.70	0.55
	Itti-Koch [9]	0.76	0.68	0.68	0.44	0.48	3.93	1.05	0.88	0.28
EMOd	ESM-Sal	**0.83**	**0.81**	0.71	0.59	0.54	2.95	1.61	0.77	1.41
	Non-ESM-Sal	**0.83**	0.80	0.71	0.61	0.54	2.76	1.68	0.82	1.35
	SALICON [14]	**0.83**	0.80	0.70	**0.64**	0.55	2.73	**1.75**	**0.72**	**1.51**
	SalGAN [27]	0.82	0.80	**0.76**	**0.64**	**0.58**	**2.63**	1.74	0.82	1.15
	SROD [12]	0.75	0.74	0.67	0.37	0.41	4.41	1.00	1.15	0.87
	BMS [10]	0.71	0.68	0.60	0.30	0.40	4.07	0.83	2.18	0.56
	GBVS [11]	0.79	0.78	0.62	0.46	0.47	3.32	1.20	0.97	1.12
	Itti-Koch [9]	0.75	0.74	0.64	0.38	0.43	4.04	0.99	1.10	0.95
Affective CAT2000	ESM-Sal	0.85	**0.84**	0.67	0.67	0.57	5.15	1.96	0.76	29.09
	Non-ESM-Sal	0.85	0.82	0.67	0.68	0.57	**4.33**	2.04	0.83	28.97
	SALICON [14]	**0.86**	0.82	0.67	**0.69**	**0.59**	4.50	**2.08**	**0.71**	**29.20**
	SalGAN [27]	0.86	0.83	**0.68**	**0.69**	0.58	5.27	2.04	0.94	28.83
	SROD [12]	0.81	0.80	0.64	0.46	0.45	6.87	1.32	1.04	28.69
	BMS [10]	0.78	0.73	0.59	0.39	0.44	5.97	1.16	1.86	28.56
	GBVS [11]	0.83	0.82	0.60	0.52	0.48	6.08	1.49	0.90	28.89
	Itti-Koch [9]	0.80	0.79	0.61	0.44	0.44	7.37	1.26	1.02	28.72

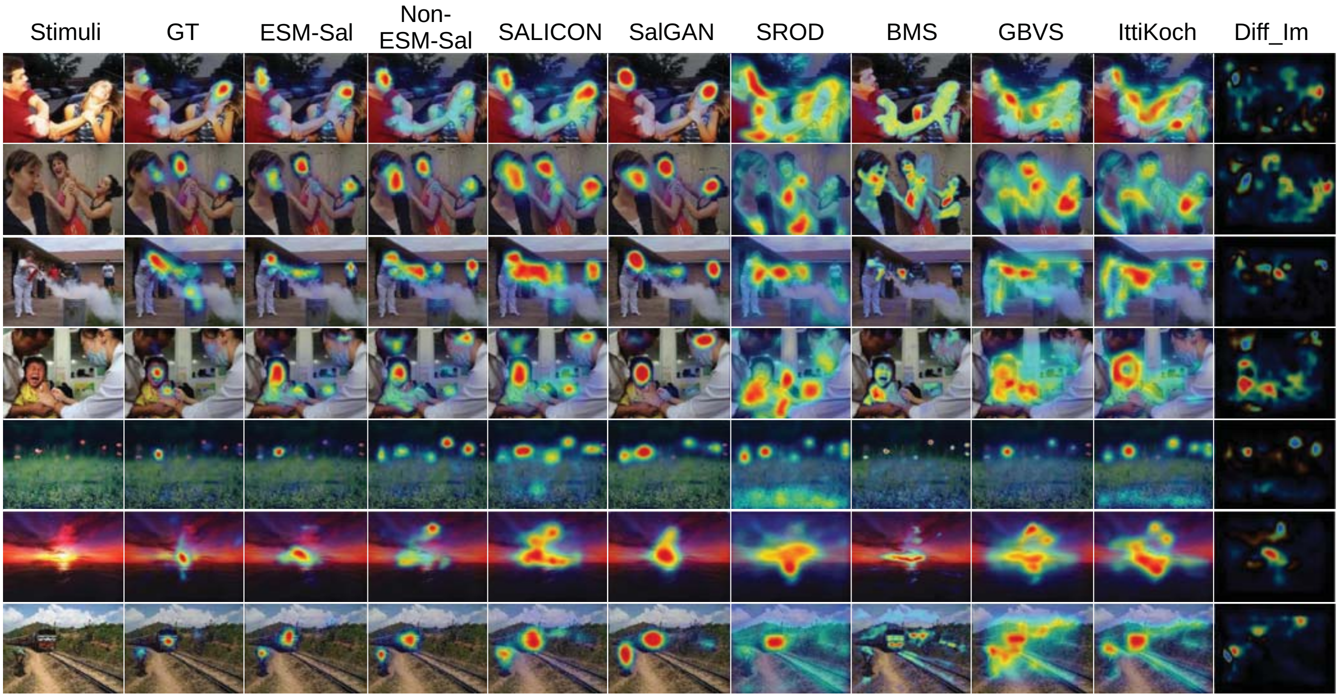

Fig. 3. Qualitative results showing the saliency map outputs from ESM-Sal and other saliency models on sample images taken from NUSEF and EMOd. The last column corresponds to the difference between the ESM-Sal and Non-ESM-Sal predicted saliency map. The redness in the Diff_Im column indicates that the saliency increases in ESM-Sal with respect to Non-ESM-Sal and the blueness of the Diff_Im column indicates decrease in saliency prediction in that region.

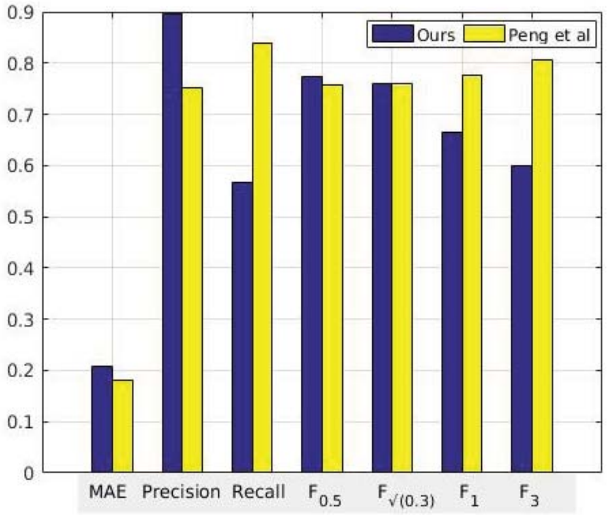

Fig. 4. ESM prediction which predicts the emotion-inducing region in an image was introduced in [8] and evaluated quantitatively by the metrics e.g. Mean Absolute Error (MAE), the F-measures. The performance of the implemented ESM prediction, developed as the saliency prediction was being trained, shows comparable performance in terms of MAE, Precision, $F_{0.5}$ and $F_{\sqrt{0.3}}$.

Non-ESM-Sal. As not all persons in an image are important which can be seen in the ground truth fixation maps, ESM-Sal is able to adjust this by minimizing the saliency values of over estimated objects and increasing the values of under estimated objects. For the first two examples, the ground truth shows that the most salient persons are those who receive the provoking action of the actors. The Non-ESM-Sal prediction, over-estimated other persons in the image, which is corrected in ESM-Sal prediction. In the third image, the salient objects based on the ground truth are the person and its action. However in Non-ESM-Sal, other persons in the scene are also considered as salient which is properly adjusted in ESM-Sal.

The fourth to sixth examples show the qualitative improvement in terms of misdetections. In the fourth stimuli, the ground truth shows that the salient regions are the face of the crying boy and the injection. The Diff_Im column shows that there is an increase in the saliency values in that region, trying to correct the misdetection. Similarly for the fifth and sixth images with the salient region being the flower and the sun, respectively, Non-ESM-Sal misdetected these objects as salient but are corrected in the ESM-Sal predictions. Additionally, those objects with over estimated saliency values are adjusted in ESM-Sal.

Finally, to ensure that the ESM prediction actually performs its task, we evaluated this using the test set of EmotionROI [8] and the following metrics, the Mean Absolute Error (MAE), Precision, Recall, the F-measures and the Precision-Recall curve used by the ESM prediction proponents. As summarized in Fig. 4 Except for the Recall, F_1 and F_3 scores, the resulting ESM prediction model has comparable performance with the model described in [8].

V. The contribution of emotion

As discussed in [2], humans tend to fixate on people in an image that are central to a depicted action, a conversation, an event or people who stand out from the crowd based on some high-level features. By extension, emotion-inducing regions stand out from the other parts of the image and tend to attract the attention of the observer. Some examples of this are shown in Figure 1 where the emotion-inducing region (indicated by yellow markers) tend to have high saliency values in the final saliency (ESM-Sal column) over the non-emotion-inducing image portion.

So how does the introduction of emotion-inducing region affect saliency prediction? During the training, the error from ESM and saliency prediction are propagated from the fifth layer of the feature detector down towards the first layer. This means that any misdetection of the affective region are corrected node by node down to the feature detectors. Ultimately, nodes which are not connected to the affective region are not activated and thus, are not developed or updated during the training sequence. This can be seen in the correspondence of the response of the resulting emotion-attentive saliency model when the predicted affective region is mapped on to the Diff_Im. It can be seen that the area which are modified in emotion-attentive model, as illustrated in Fig. 5 are strongly correlated to the predicted emotional region. That is, the last column shows that most of the saliency values correction in ESM-Sal are concentrated inside the predicted emotional region.

VI. Conclusion

As saliency prediction performance starts to saturate, researchers start to look for higher-level concept that would allow better saliency prediction to approach human-level performance. In this work, we explored the use of emotion to improve saliency prediction. We use the emotion-inducing region as the image attribute to facilitate the emotion awareness in saliency prediction. We trained the VGG feature detector, cascaded with two inception blocks, simultaneously with the Emotion Stimuli Map (ESM) prediction model which essentially increase the development of neurons both involved in saliency prediction and ESM prediction, faster and better. Using three saliency datasets with emotion-rich images, quantitative metrics show that our proposed emotion-attentive saliency model (ESM-Sal) is at par with the current state-of-the-art complex saliency models.

With respect to a similar model ”unaware” of the emotion-inducing region (Non-ESM-Sal), our ”emotionally” developed model has improved in terms of mis-detection and relative saliency prediction, which can be observed in better IG and KL metrics and in qualitative comparison. Analysis of the output saliency map and the output ESM shows that the difference between the ESM-Sal and Non-ESM-Sal models is mainly due to the activation nodes inside the emotion-inducing region, contributing to better focus on salient regions, as manifested by the amount of adjustment in the emotional and non-emotional region.

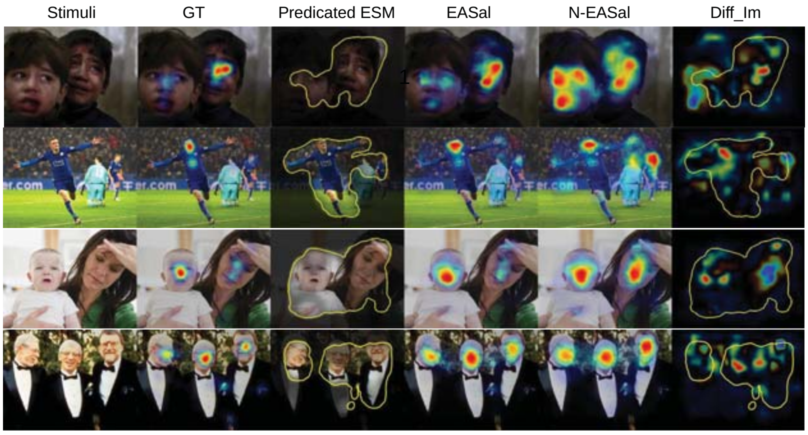

Fig. 5. Close comparison of the predicted ESM and the saliency prediction shows strong correspondence between the predicted emotional region enclosed by yellow curve and the changes in the saliency map (last column). It can be seen in the Diff_Im column that there is a strong agreement between the location of the those regions with corrected saliency values and the emotional region. This agreement suggests that the neurons of the feature detector which are directly related to the identification of the emotional region are adjusted and corrected during the training iteration, allowing better saliency prediction.

REFERENCES

[1] Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba, "Mit saliency benchmark."

[2] Z. Bylinskii, A. Recasens, A. Borji, A. Oliva, A. Torralba, and F. Durand, "Where should saliency models look next?" in *ECCV*. Springer, 2016, pp. 809–824.

[3] S. Frintrop, S. Rome, and H. Christensen, "Computational visual attention systems and their cognitive foundations: A survey," 2010.

[4] D. Lane, Richard, M.-L. Chua, Phyllis, and J. Dolan, Raymond, "Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures," *Neuropsychologia*, vol. 37, pp. 989–997, 1999.

[5] P. Vuilleumier, "How brains beware: neural mechanisms of emotional attention," *Trends in Cognitive Sciences*, vol. 9, no. 12, pp. 585–594, 2005.

[6] P. Bradley, Brendan, K. Mogg, N. Millar, C. Bonham-Carter, E. Fergusson, J. Jenkins, and M. Parr, "Attentional biases for emotional faces," *Cognition and Emotion*, vol. 11, no. 1, pp. 25–42, 1997.

[7] D. Ren, P. Wang, H. Qiao, and S. Zheng, "A biologically inspired model of emotion eliciting from visual stimuli," *Neurocomputing*, vol. 121, pp. 328–336, 2013.

[8] K.-C. Peng, A. Sadovnik, A. Gallagher, and T. Chen, "Where do emotions come from? predicting the emotion stimuli map," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016.

[9] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[10] J. Zhang and S. Sclaroff, "Saliency detection: A boolean map approach," in *2013 IEEE International Conference on Computer Vision*, Dec 2013, pp. 153–160.

[11] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. C. Platt, and T. Hoffman, Eds. MIT Press, 2007, pp. 545–552.

[12] C. Chuanbo, T. He, L. Zehua, L. Hu, S. Jun, and S. Mudar, "Saliency modeling via outlier detection," *Journal of Electronic Imaging*, vol. 23, no. 5, 2014.

[13] M. Treisman, Anne and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97 – 136, 1980.

[14] X. Huang, C. Shen, X. Boix, and Q. Zhao, "Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks," in *Computer Vision (ICCV), 2015 IEEE International Conference on*. Santiago, Chile: IEEE, 2016, pp. 262–270.

[15] S. S. Kruthiventi, Srinivas, K. Ayush, and R. V. Babu, "Deefix: A fully convolutional neural network for predicting human eye fixation," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4446–4456, 2017.

[16] M. Kümmerer, L. Theis, and M. Bethge. DeepGaze I: Boosting saliency prediction with feature maps trained on ImageNet.

[17] Kümmerer, Matthias and Theis, Lucas and Bethge, Matthias. DeepGaze II: Reading fixations from deep features trained on object recognition.

[18] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara, "Predicting human eye fixations via an LSTM-based saliency attentive model," vol. 27, no. 10, pp. 5142 – 5154, 2018.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[20] A. Krizhevsky, I. Sutskeyer, and E. Hinton, Geoffrey, "Imagenet classification with deep convolutional neural network," in *NIPS2012 Proceedings of the 25th International Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.

[21] T. Brosh, K. Scherer, D. Grandjean, and D. Sander, "The impact of emotion on perception, attention, memory and decision-making," 2013.

[22] K. R. Mickley Steinmetz and E. A. Kensinger, "The emotion-induced memory trade-off: more than effect or overt attention?" vol. 41, pp. 69–81, 2013.

[23] H. Okon-Singer, J. Mehnert, J. Hoyer, L. Hellrung, H. L. Schaare, J. Dukart, and A. Villringer, "Neural control of vascular reactions: Impact of emotion and attention," vol. 34, no. 12, pp. 4251–4259, 2014.

[24] G. Hajcak, A. MacNamara, D. Foti, J. Ferri, and A. Keil, "The dynamic allocation of attention to emotion: Simultaneous and independent evidence from the late positive potential and steady state visual evoked potentials," vol. 92, no. 3, pp. 447–455, 2013.

[25] M. Sun, J. Yang, K. Wang, and H. Shen, "Discovering affective regions in deep convolutional neural networks for visual sentiment prediction," 07 2016, pp. 1–6.

[26] H. Zheng, T. Chen, and J. Luo, "When saliency meets sentiment: Understanding how image content invokes emotion and sentiment," *AAAI*, 2016.

[27] J. Pan, C. Canton, K. McGuinness, N. E. O'Connor, J. Torres, E. Sayrol, and X. a. Giro-i Nieto, "Salgan: Visual saliency prediction with generative adversarial networks," in *arXiv*, January 2017.

[28] S. Ramanathan, H. Katti, N. Sebe, M. Kankanhalli, and T.-S. Chua, "An eye fixation database for saliency detection in images," in *ECCV 2010*, Crete, Greece, 2010.

[29] J. Lang, Peter, M. Bradley, Margaret, and B. N. Cuthbert, "International affective picture system (iaps): Affective ratings of pictures and instruction manual," University of Florida, Tech. Rep. Technical Report A-8, 2008.

[30] S. Fan, M. Jiang, J. Xu, B. Koenig, Y. Cheng, M. Kankanhalli, and Q. Zhao, "A correlational study between human attention and high-level image perception," *Journal of Vision*, vol. 17, no. 10, pp. 705–705, 2017.

[31] A. Borji and L. Itti, "Cat2000: A large scale fixation dataset for boosting saliency research," *CVPR 2015 workshop on "Future of Datasets"*, 2015, arXiv preprint arXiv:1505.03581.