HEAT-MAP BASED EMOTION AND FACE RECOGNITION FROM THERMAL IMAGES

A Thesis

Presented to

The Faculty of the Department of Computer Science

Sam Houston State University

In Partial Fulfillment

of the Requirements for the Degree of

Master of Science

---

by

Burak Ilikci

May, 2019

HEAT-MAP BASED EMOTION AND FACE RECOGNITION FROM THERMAL IMAGES

by

Burak Ilikci

APPROVED:

Qingzhong Liu, PhD
Thesis Director

Mingkui Wei, PhD
Committee Member

Hyuk Cho, PhD
Committee Member

John B. Pascarella, PhD
Dean, College of Science and Engineering Technology

ABSTRACT

Ilikci, Burak , *Heat-map based emotion and face recognition from thermal images*. Master of Science (Computing and Information Science), May, 2019, Sam Houston State University, Huntsville, Texas.

Nowadays, emotion recognition has become a feasible problem with implementation of Convolutional Neural Networks in Computer Vision domain. However, credibility of emotion recognition from daily images or videos is not enough. As people can easily mimic emotions one after another and fooling the trained models, a different approach should be taken into consideration. Thermal cameras would be a suitable way to develop more credible emotion recognition models. Heat-map of faces proved hinting emotions before, and it is not easy to fool the models trained from thermal heat-maps as it visualizes state of the body's heat. In this research a method is adapted for training a model for recognizing emotions from thermal heat-mapped cameras with a fast detection algorithm -YOLOv3-. With this method the main aim is to detecting emotions from a given picture which taken from thermal cameras.

KEY WORDS: Thermal images, Emotion recognition, Convolutional Neural Network, Eigen-space Method, YOLOv3

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

CHAPTER I

Introduction

With the emergence of Convolutional Neural Networks, image/video processing and recognition of various categories of a material became more feasible. Automated emotion recognition from humans relies as one of the most fascinating problems that seems possible to achieve within the domain of Deep Learning. As machines becoming more humanoid entities day-by-day, detecting emotions autonomously stands as a crucial challenge for researches because emotion detection is an essential skill for an 'intelligent' being and we can make machines work on this problem as a person does.

Detecting emotions is not just technical challenge for programming, the success also relies within the quality lightning conditions in the photo. To tackle this problem, thermal images comes as a reliable technique. Infrared cameras are not sensitive to light conditions. Also, since emotions is related with neurological effect of a body's own, heat maps taken from infrared cameras can be a great clue to detect emotions. The other advantage of thermal images are faking states of emotion with mimics, however, it is not comparatively easy to fake expressions with infrared cameras. It was examined that change of emotions that caused by external sounds, objects and actions causes plummet or increase in heat of a person's forehead, mouth and cheek area[1]. Therefore, thermal imaging is more reliable method to estimate human emotions because it is contract free, non-invasive and free from startling the subject in case getting abnormal results in an experiment session[2]. In addition, it is important to point out that facial thermal images do not depend on skin or color of a subject[3].

Infrared spectrum is divided into four sub-bands: near IR (NIR), short wave IR (SWIR), medium wave IR (MWIR) and long wave IR (LWIR). Most of the heat energy radiated from infrared spectrum is LWIR sub-band and after LWIR, MWIR does also radiates a significant amount of heat-wave. That is why LWIR and MWIR sub-band supported cameras is used with most of the work in the literature[4].

Robustness of thermal images in detecting human emotions can bring great benefits for applications which needs correct interpretation of human emotions. To name few areas where emotion detection is useful; socially aware systems[5,6], socially skilled robots[7], frustration analysis of students in e-learning[8], tracing a patient's pain analysis in clinical environments[9, 10, 11], detecting if a person is lying during interrogation or interview process[12] may be given as examples of benefits of emotion detection with thermal images[13].

Structure of the paper is; section II focuses on the techniques implemented in mentioned works, section III describes the methodology to process taken image or video from a thermal camera to detect emotions with pre-trained convolutional architecture model, section IV portrays the timeline of process of the research.

CHAPTER II

Background

As a recognition technique for detecting the faces from a given image or video, Eigenface-based approach and Principal Component Analysis (PCA) lies as crucial ones that helps extracting the features of images to mathematical entities[14]. In order to detect the heat levels taken from infrared cameras to detect emotion, it is important to divide the image into sub-space parts and examine the differences by those parts is helpful to detect emotions. Those sub-parts of an image are called Region of Interests (ROIs) in the literature.

Nguyen et al.[15] proposed a method of automation of emotion recognition by proposing a temperature space method to correct an alien object's effect on eye region, and later using PCA, Eigen-space method based on class-features EMC, and PCA-EMC method to classify emotions from the images which was made more suitable for mathematical operations to obtain a result. They used Kotani Thermal Facial Emotion (KTFE) database[16]. As a result, Nguyen et al. got maximum of %66.19 accuracy with PCA-EMC with subject having an eye glass on and %73.96 accuracy with PCA after the eye glass was taken off.

Another method was proposed by Basu et al.[17] which divides the thermal image to 6 specific sub-regions (forehead, eyes, left cheek, right cheek, nose and mouth) and uses Hu's moment invariants to calculate feature vectors from the six facial paths to create ROI for emotions. Then, statistical feature of each ROI is computed and fused with moment invariant and later Multi-class Support Vector Machine is used as a classification tool and a suitable kernel function is chosen to maximize the inter class distance and to minimize overlapping between classes. Basu et al. used KTFE database. Overall, they have achieved average of %87.5 accuracy.

Next method considered for this research uses Eigen face technique which reduces dimension of images and helps dealing with large dataset of face images. Later, PCA derives the weights from original images and ADA Boost algorithm is applied in order to simplify and reduce the number of weights to boost calculation. Scale Invariant Feature Transform (SIFT) algorithm extracts the eyes, nose and lips for detailed description of the image and Gray Level Co-occurrence Matrix (GLCM) holds the numbers of pixel and position details. Finally, obtained gray images from previous steps faces applied to Feed Forward Neural Networks (FFNN) to train the model[18]. The final accuracy for 80 pictures is above %94 for each emotions and results are better than Support Vector Machines.

Another research by Irving et al.[19] implements a smart thermal system and examines emotion detection while emotion changes are happening in a time span. Subjects were placed in an environment while they were getting recorded by a thermal camera and they were under surveillance by an ‘expert’. The emergence of emotions was planned to take place in subject with the help of pre-selected videos with the help of experts from the Psychology area because it was stated that videos are one of the most effective tools to stimulate emotions[20]. Technique to detect emotions as follows; first thermogram is taken from subject at the first phase and ROIs are analyzed as the previous mentioned works done. ROIs are examined by a fuzzy logic system; 4 inputs are considered: forehead, cheeks, nose and maxillary. After, when an emotion is forming, subject’s thermogram is taken again and analysis of ROIs are calculated one more time and 4 inputs (forehead, cheeks, nose, maxillary) passes new temperature levels to calibration step. Calibration step divides and compares the two pictures ROIs and divides the phases into 3 groups; low means temperature has decreased, normal means temperature stayed still, and high means temperature has risen. Based on change of temperatures, outputs are created for each ROI stating their status (increased or decreased). Finally, the classifier they implemented diagnoses one of the five emotions (joy, disgust, fear, anger, sadness) by using top-down hierarchical classifier which considers the temperature differences by obeying the emotion rules. Overall, the accuracy for 8 male subjects was %90.3 and 17 female subjects was %89.5 resulting %89.9 overall accuracy.

This research proposes to apply You Only Look Once v3 (YOLO v3)[21] algorithm for the thermal emotion detection. YOLOv3 works with a custom deep architecture called Darknet and it has 53 convolutional layer network trained on one of the image database platforms. It uses logistic regression and comes up with an objectness score for each bounding box which is predicted by the network.

CHAPTER III

Methodology

The generated pictures will be applied to YOLOv3 [21] for training. The network contains 53 convolutional layers as shown in Figure 1.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure 1. Architecture of Darknet 53 which was implemented for the use of YOLOv3 [21].

The system predicts bounding boxes using dimension clusters as anchor boxes. Then, anchor boxes determine the objectness score for the object and displays what is the object. The mentality of bounding boxes is shown in figure 2[23].

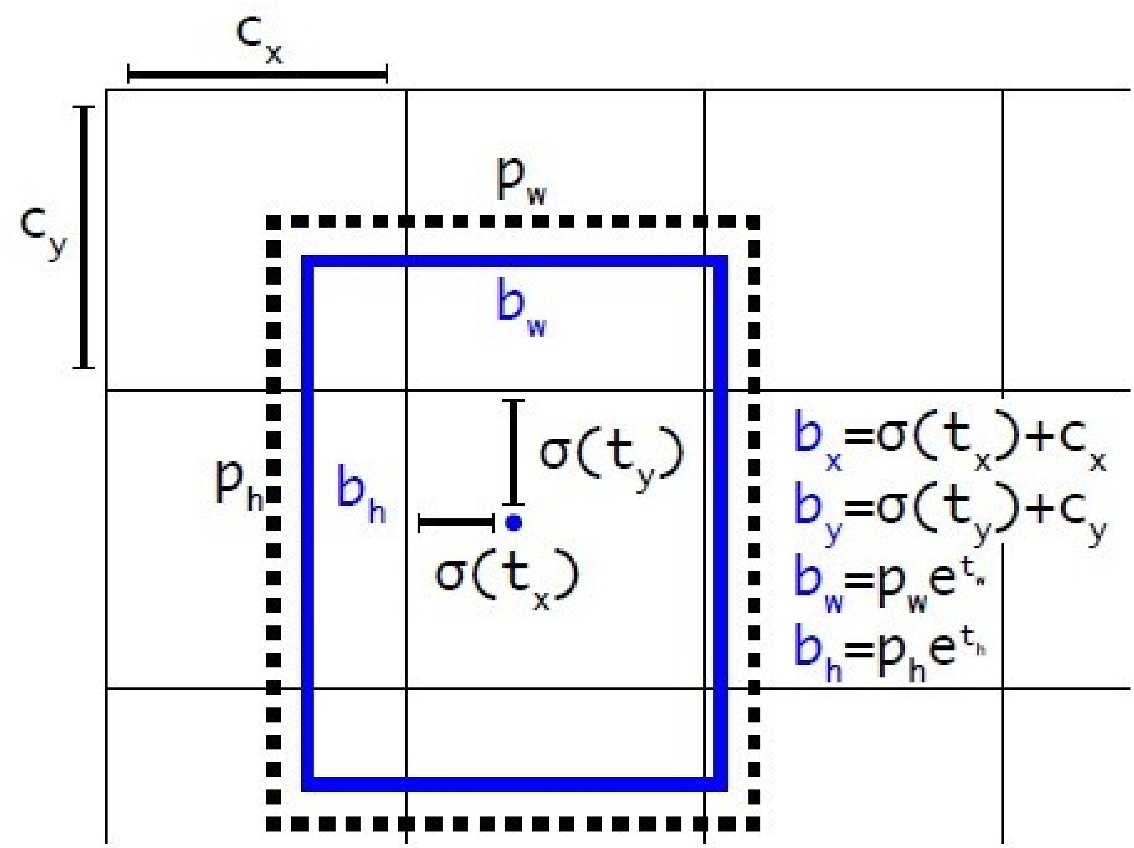


Figure 2. The width and height of the box as offsets from cluster centroids. Center coordinates of the box is predicted by using sigmoid function[23].

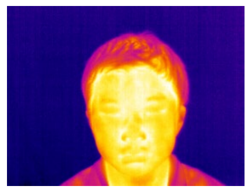
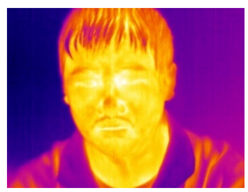
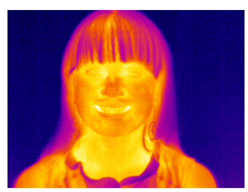
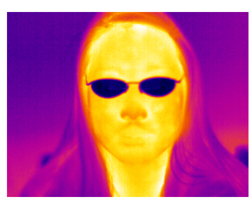
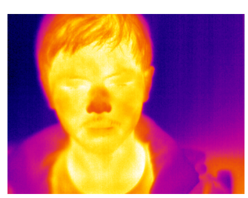
CHAPTER IV

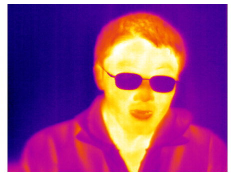
Dataset

Facilities for conducting an emotion recognition research for deep convolutional networks is scarce. As a first step searching for suitable dataset which contains posed pictures of individuals with sorted by classification of emotions took a long time. Unfortunately, there are not many data sets of faces taken from LWIR and MWIR sub-band supported cameras. As one of the few, NVIE (Natural Visible and Infrared Facial Expression) Database from University of Science and Technology of China provides an intensive thermal face database[22]. The database contains pictures of more than a hundred subjects with posed and spontaneous expressions. Few different factors such as illuminating from three different perspective and images of individuals with glasses and non-glasses were considered. For classification purposes, 13 points in each individual's picture are pointed manually. With manually marked points PCA and PCA+LDA techniques were used in order to examine the relationship with statistical analysis between facial temperature and emotion. For training the model, only posed infrared images were used from NVIE Database. Posed database contains a total of 3452 images which contains seven emotions; anger, disgust, fear, happiness, neutral, sadness and surprise. The number of each emotion and one example for each emotion are shown in the Table 1.

Table 1

Distribution of types of classified emotions in the dataset with an example

Type of Emotion	Example	Count
Anger		549
Disgust		535
Fear		550
Happiness		550
Neutral		206
Sadness		532

Type of Emotion	Example	Count
Surprise		530
	Total:	3452

YOLO

YOLO algorithm divides a picture into SxS grid. Each grid is responsible for detecting an object if the center of it encounters to that specific grid. After the prediction is made from a grid, bounding boxes label the object with their confidence score. However, this method may result having more than one bounding box were put by model for an object. In order to eliminate multiple labels to one on an object, predicted score of object gets multiplied with Intersection over union scores between bounding boxes. Highest scored bounding box remains on object and others are eliminated.

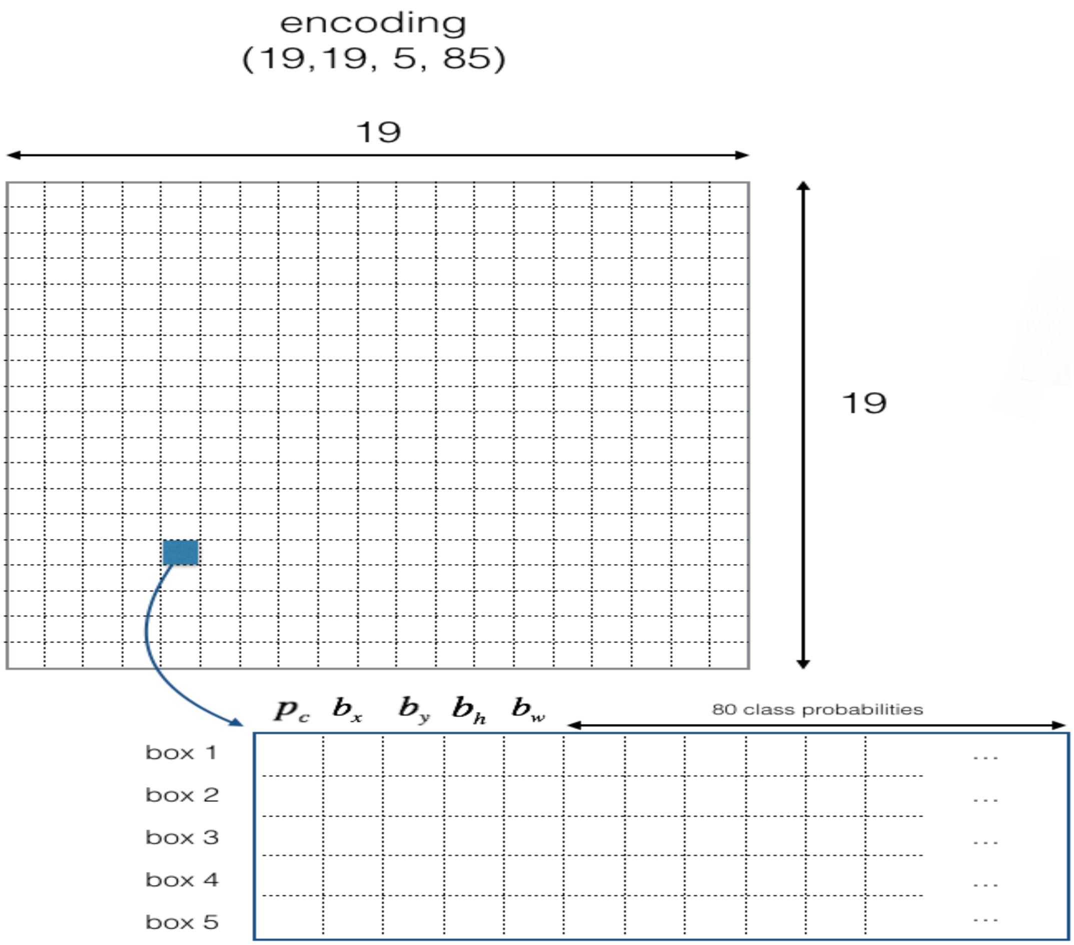


Figure 3. Visualization of grids with an example of 19x19 with 5 anchor boxes and 80 class probabilities

YOLO requires a text file for image with labels prepared on each object for training the model. In order to strengthen the model, each image in the selected database labeled one by one. As a result, a text file was generated with following format:

<class of object> <x coordinate of label> <y coordinate of label> <width of label> <height of label>[24].

YOLO is open-sourced and publicly available in order to use the neural architecture to train a model. Although NVIE database contains information of coordinates of specific marked points in each picture, due to required format difference, information of bounding box coordinates of emotions for each picture created from scratch and included into the source code.

Results

Pre-trained convolutional weights are used as initial weights for the model. These weights are produced by Darknet-53 model and trained on Imagenet[25]. These weights provide a great start for training emotion detection model. Model was trained for 9000 iterations. In each iteration batch size was 64 and subdivision was 16. Threshold for detecting emotions set to the confidence of 25 percent. After each thousands iterations, weights were saved for comparison purposes.

Average precision of each classes began from average of 25 percentages and after 9000 iterations it raised up to 90 percent for each classes except fear and neutral. Mean average precision in last iteration is 92.72 percent. As shown in the Figure 5 the model is really good at detecting happiness. It is also good in detecting surprise and disgust. Other emotions; anger, fear, neutral and sadness, are the ones that the model was having bad precision scores. That is correlational with having multiple detections for an image. As can shapes of specific locations such as position of eyebrows, visibility of lips, etc… can be similar during different emotions. Average IoU (Intersection over Union) also indicates that with a final score of 71.65 percent.

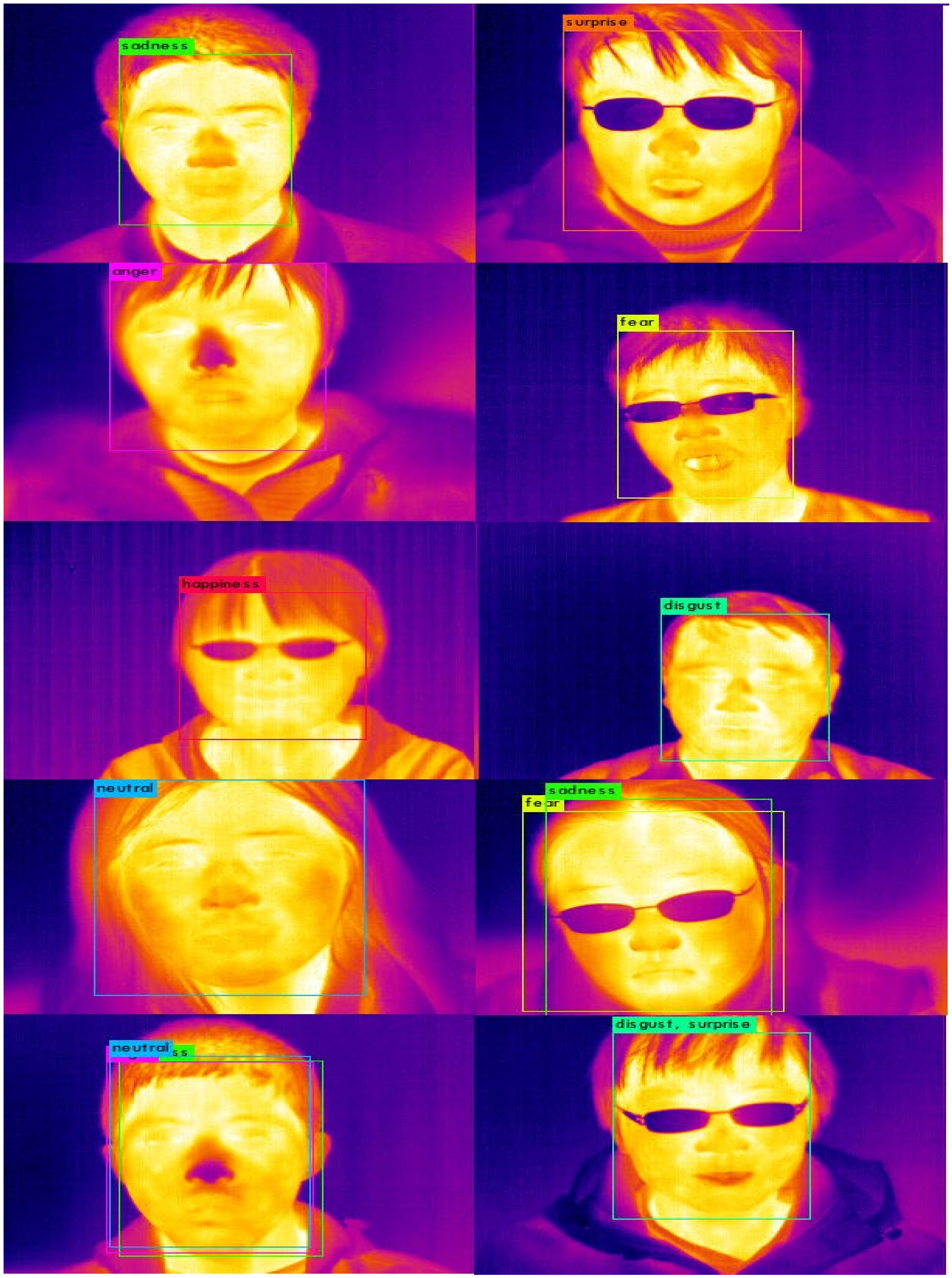


Figure 4. Examples of detections. The model predicted more than one emotions in some cases which shows similarity of face positions may lead more than one detection.

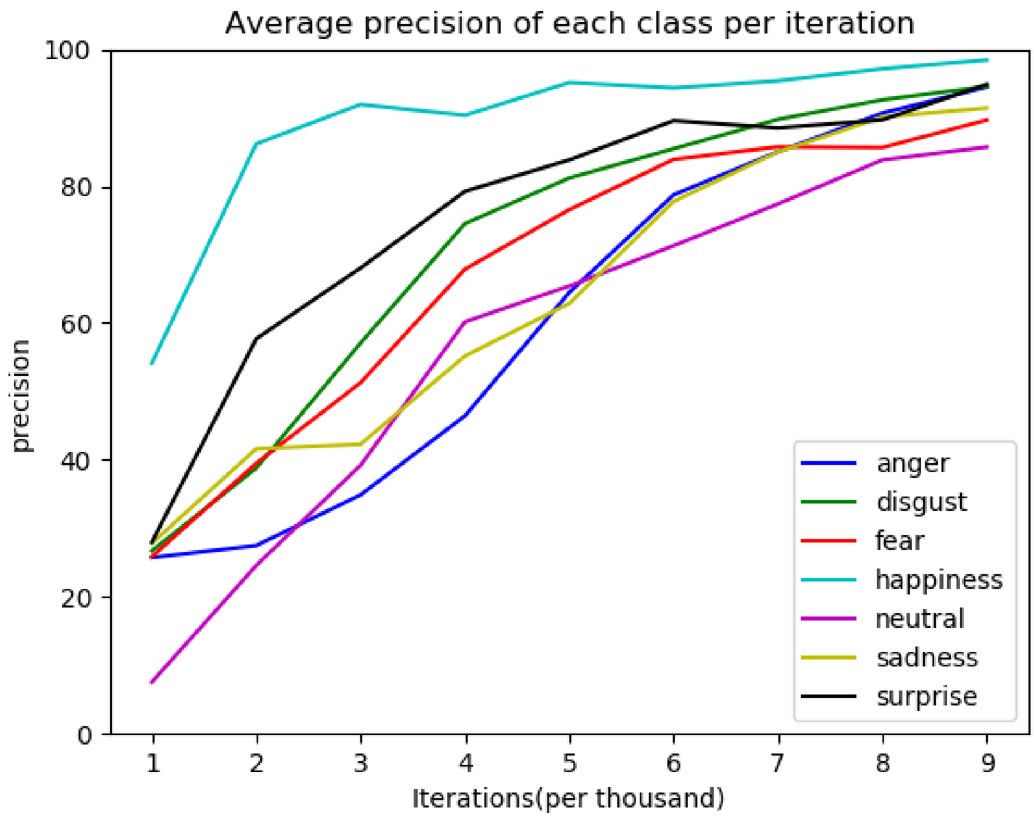


*Figure 5.*Average precision of each detection per iteration.

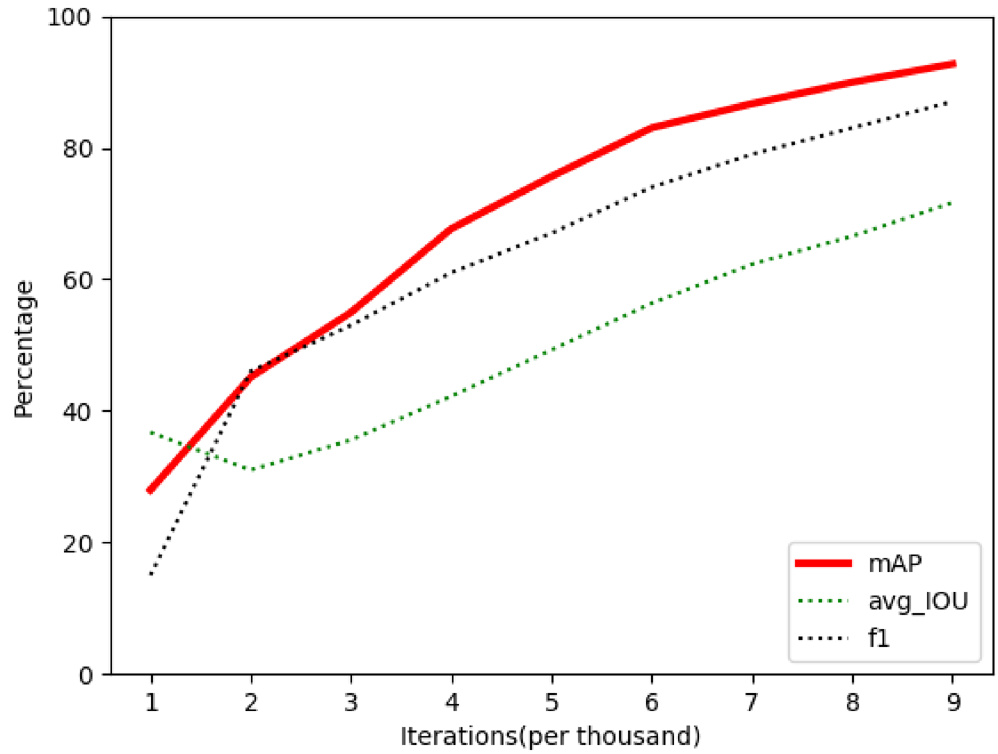


Figure 6. Mean Average Point(mAP), Average Intersection Over Union(IoU) and F1 scores per iteration.

Comparison

For the sake of comparison emotion detection from NVIE Dataset also trained with DenseNet-201[26] and ResNet-152[27]. Average IoU and mean average precision scores for each thousand iterations up to 9000 for YoloV3, DenseNet-201 and ResNet-152 are shown in the Table 2. In addition, graphs also included in Figure 7 and Figure 8.

Table 2

mAP and IoU scores of YOLO, DenseNET and ResNET in each iteration up to 9000

	YOLO		DenseNET		ResNet	
Number of Iterations	mAP	IoU	mAP	IoU	mAP	IoU
1000	%27.97	%36.73	%43.88	%33.37	%43.59	%29.35
2000	%45.12	%31	%56.6	%35.91	%58.08	%38.01
3000	%54.96	%35.56	%73.84	%46.25	%67.28	%40.64
4000	%67.67	%42.22	%82.97	%50.95	%78.91	%47.86
5000	%75.63	%49.26	%87.18	%58.4	%83.63	%54.55
6000	%83.01	%56.34	%92.67	%66.23	%89.13	%62.28
7000	%86.68	%62.29	%93.65	%71.87	%90.76	%67.84
8000	%89.94	%66.54	%93.61	%73.44	%92.65	%71.8
9000	%92.72	%71.65	%96.46	%74.55	%94.2	%73.31

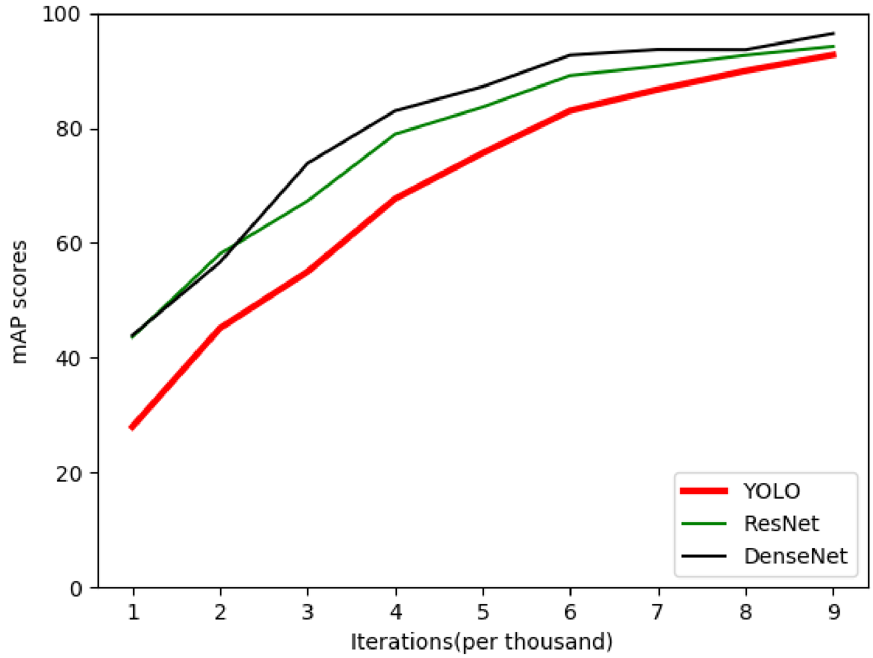


Figure 7. Mean Average Precision scores of YOLO, DenseNET and ResNET.

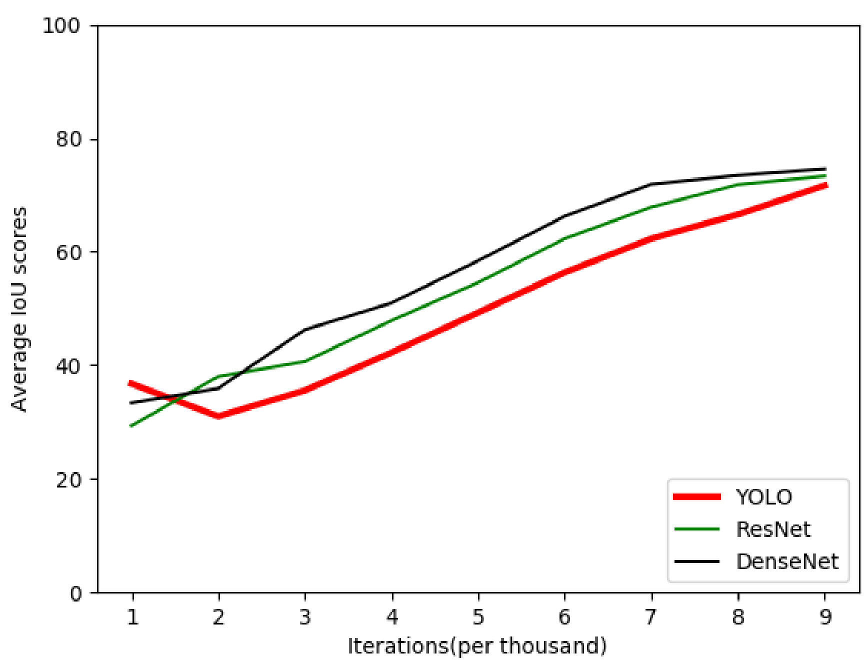


Figure 8. Average Intersection Over Union scores of YOLO, DenseNET and ResNET.

As also mentioned in YOLOv3[21], DenseNet and ResNet perform better on mAP calculations because that models are deeper in comparison with Darknet 53's architecture. However, this advantage of DenseNet and ResNet costs more time as it becomes longer to train these networks. On the other side when it comes to detection times, DenseNet and ResNet are slower than YOLOv3.

In this experiment, it can be seen from the figures and table in this section, DenseNet and ResNet jumped quickly on measurements with DenseNet reaching %90 mAP on 6000 iterations and ResNet on 7000 iterations. However, if the curve of YOLO taken into consideration in Figure 7, it did not stall in last 2000 iterations and caught up other algorithms mAP scores. In coherence with YOLOv3's research the detection times were faster compared to other algorithms.

Weaknesses

Model performs greatly on the dataset. However, the advantage of the dataset becomes a hurdle when it comes to test the model with images from outside. Since database is homogenous the model does not reflect on other images as it does to the dataset. Model tends to label an image with 2 emotions more often as far as I observed. Also, model rarely detect emotions in outsourced image with two contradictory states such as happiness-sadness or disgust-surprise etc…

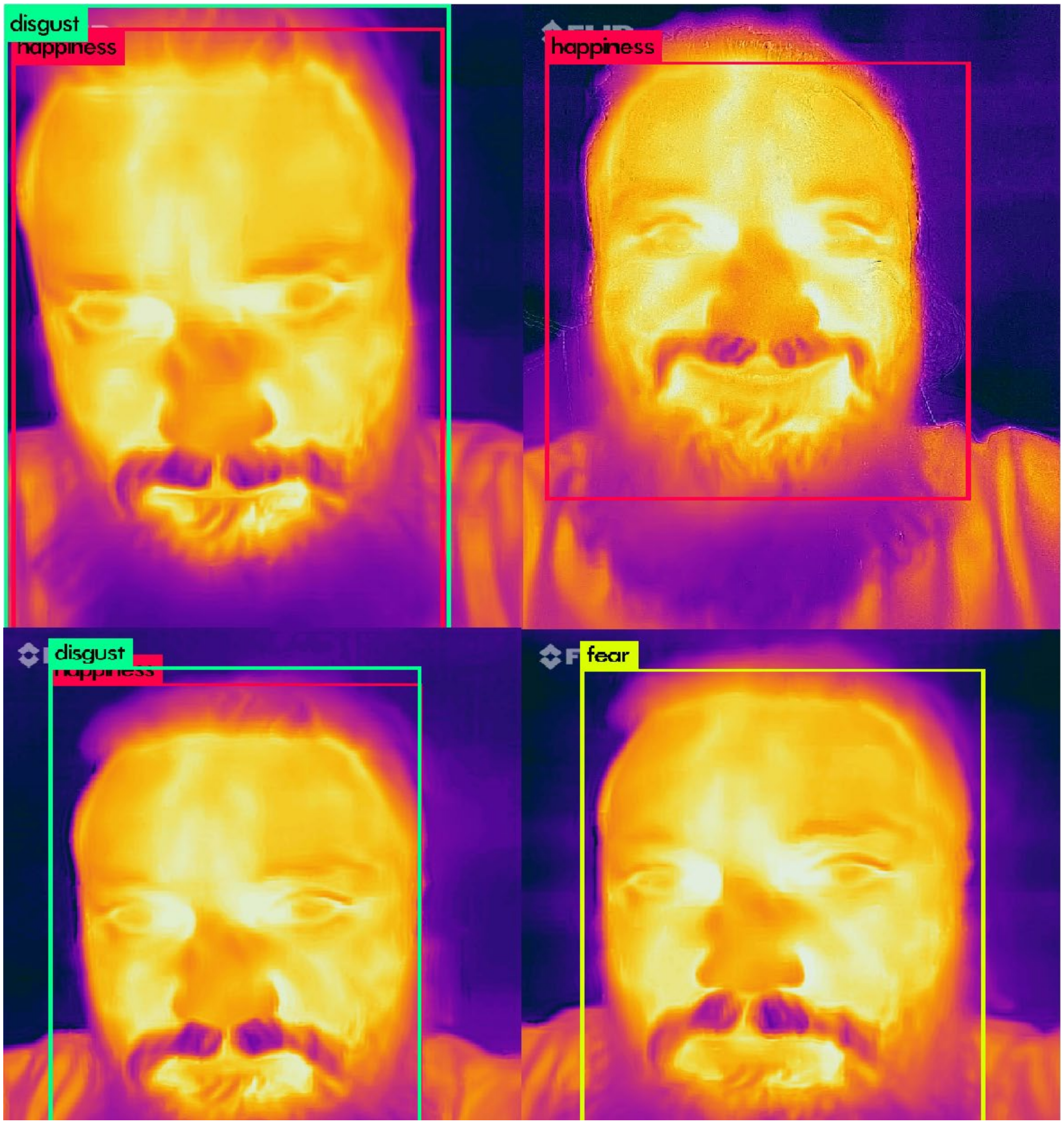


Figure 9. Testing model with outsourced images.

Conclusion & Future Work

YOLOv3 is such a great algorithm to detect objects. The trained model in this research shows that YOLO algorithm and Darknet architecture can potentially be used for detecting emotions. The problem is lack of thermal databases in academia. For future work there could be an attempt to build a dataset with multiple people posing in a picture which potentially would be more reliable approach to be tested in daily life for tasks to help in psychology and internal security. Emotions play a key role in detecting an individual's psychology. As for security people who are going to commit crime or disturbance in the public might be detected from their unusual heat levels.

REFERENCES

[1] E. Engelhaupt, "Your fear is written all over your face, in heat," Science News, 28-Mar-2014. [Online]. Available: https://www.sciencenews.org/blog/gory-details/your-fear-written-all-over-your-face-heat.

[2] A. Topalidou, A. Nazmin, "Infrared emotions and behaviours: Thermal imaging in psychology," International Journal of Prenatal and Life Sciences, Vol 1, No 01, p. 65-70, Aug 2017.

[3] J. Eom, J. Sohn, "Emotion Recognition using Facial Thermal Images," Journal of Ergonomics of Society of Korea, Vol 31, Issue 3, p.427-435, 2012.

[4] R. S. Ghiass, O. Arandjelović, A. Bendada, and X. Maldague, "Infrared face recognition: A comprehensive review of methodologies and databases," Pattern Recognition, vol. 47, no. 9, pp. 2807–2824, 2014.

[5] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," IVC, vol. 27, no. 12, pp. 1743–1759, 2009.

[6] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, A. Gainer, and L.-P. Morency, "A virtual human interviewer for healthcare decision support." AAMAS, 2014.

[7] H. Ishiguro, T. Ono, M. Imai, T. Maeda, T. Kanda, and R. Nakatsu, "Robovie: an interactive humanoid robot," Industrial robot: An international journal, vol. 28, no. 6, pp. 498–504, 2001.

[8] A. Kapoor, W. Burleson, and R. W. Picard, "Automatic prediction of frustration," IJHCS, vol. 65, no. 8, pp. 724–736, 2007.

[9] P. Lucey, J. F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, and K. M. Prkachin, “Automatically detecting pain in video through facial action units,” SMC-B, vol. 41, no. 3, pp. 664–674, 2011.

[10] S. Kaltwang, O. Rudovic, and M. Pantic, “Continuous pain intensity estimation from facial expressions,” ISVC, pp. 368–377, 2012.

[11] R. Irani, K. Nasrollahi, M. O. Simon, C. A. Corneanu, S. Escalera, C. Bahnsen, D. H. Lundtoft, T. B. Moeslund, T. L. Pedersen, M.-L. Klitgaard et al., “Spatiotemporal analysis of rgb-dt facial images for multimodal pain level recognition,” CVPR Workshops, 2015.

[12] A. Ryan, J. F. Cohn, S. Lucey, J. Saragih, P. Lucey, F. D. la Torre, and A. Ross, “Automated facial expression recognition system,” in ICCST, 2009.

[13] C. A. Corneanu, M. O. Simon, J. F. Cohn, S. E. Guerrero, "Survey on rgb 3d thermal and multimodal approaches for facial expression recognition: History trends and affect-related applications", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1548-1568, 2016.

[14] P. Navarrete and J. Ruiz-Del-Solar, “Comparative Study between Different Eigenspace- Based Approaches for Face Recognition,” Advances in Soft Computing — AFSS 2002 Lecture Notes in Computer Science, pp. 178–184, 2002.

[15] H. Nguyen, K. Kotani, F. Chen, and B. Le, “Estimation of human emotions using thermal facial information,” Fifth International Conference on Graphic and Image Processing (ICGIP 2013), Oct. 2014.

[16] H. Nguyen, K. Kotani, F. Chen, and B. Le, “A Thermal Facial Emotion Database and Its Analysis,” Image and Video Technology Lecture Notes in Computer Science, pp. 397–408, 2014.

[17] A. Basu, A. Routray, S. Shit, and A. K. Deb, “Human emotion recognition from facial thermal image based on fused statistical feature and multi-class SVM,” 2015 Annual IEEE India Conference (INDICON), 2015.

[18] M. S. Priya, G. M. K. Nawaz, “Modified emotion recognition system to study the emotion cues through thermal facial analysis,” Biomedical Research vol. 28, Issue 20, pp. 1-6, 2017.

[19] I. A. Cruz-Albarran, J. P. Benitez-Rangel, R. A. Osornio-Rios, and L. A. Morales-Hernandez, “Human emotions detection based on a smart-thermal system of thermographic images,” Infrared Physics & Technology, vol. 81, pp. 250–261, 2017.

[20] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, “Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers,” Cognition & Emotion, vol. 24, no. 7, pp. 1153–1172, 2010.

[21] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” arXiv, 2018.

[22] S, Wang, Z. Liu, S. Ly, Y. Lv, G. Wu, P. Peng, F. Chen, X. Wang, “A Natural Visible and Infrared Facial Expression Database for Expression Recognition and Emotion Inference,” IEEE Transactions on Multimedia, vol. 12, no. 7, pp. 682-691, 2010.

[23] J. Redmon and A. Farhadi. “Yolo9000: Better, Faster, Stronger,” arXiv, 2016.

[24] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” arXiv, 2016.

[25] A. Berg, J. Deng, and L. Fei-Fei, “Large scale visual recognition challenge 2010,” www.imagenet.org/challenges, 2010.

[26] G. Huang, Z. Liu, K. Q. Weinberger, and L. Maaten, “Densely connected convolutional networks,” In CVPR, 2017.

[27] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” In Proceedings of CVPR, pages 770–778, 2016, arxiv.org/abs/1512.03385.

VITA

Burak Ilikci

EDUCATION

Sam Houston State University (Huntsville, Texas), *Master's Degree, Computing and Information Science*

Sam Houston State University (Huntsville, Texas), *Bachelor's Degree, Computer Software Engineering Technology,* Graduation:2017

Firat University (Elazig, TURKEY), *Bachelor's Degree, Software Engineering,* Graduation:2015

EXPERIENCE

Graduate Assistant, September '17 – Now, Department of Computer Science, Sam Houston State University (Huntsville, Texas)

INVOLVEMENTS

Turkish Student Organization Founder / President, November '16 – September '17

Graduate Advisor, September '17 - Now

International Student Organization President, September '16 – September '17