Classification of facial expression dependent on lip segmentation

Reza Mohammadi Tamanani
School of Engineering
University of Guelph
Guelph, Canada
rmoham05@uoguelph.ca

Timothy Wong
School of Engineering
University of Guelph
Guelph, Canada
twong05@uoguelph.ca

***Abstract*—Extensive development into facial expression recognition techniques is currently being done due to its wide range in applications including lie detection and marketing. A convolutional neural network (CNN) classified images of happy and sad facial expressions taken from the Cohn-Kanade dataset. Because a facial expression can generally be determined by the shape of an individual's mouth, the model was trained on multiple variations of mouth images which include cropped mouth images, fine grained and spectral static saliency image maps, and mouth images automatically segmented by level set and thresholded saliency maps. The automatic segmentation methods performed very poorly, achieving a maximum dice coefficient of 47% by the thresholded spectral saliency map. The CNN classifier had a high accuracy, achieving a maximum prediction accuracy of 98% by the model trained on the fine grained saliency images. Classifiers trained using segmented images achieved an accuracy between 90% to 93%. Because the size of the dataset was 400 images, data augmentation that performed a combination of geometric transforms was used to artificially increase the size of the dataset. The CNN classifier using the original images achieved an accuracy of 85% using data augmentation, while almost models that only used mouth features achieved an accuracy of at least 90%. This indicates the potential of using exclusively mouth features to perform facial recognition in less controlled images.**

***Index Terms*—facial expression recognition, segmentation, mouth, classification, CNN**

I. Introduction

Facial expression recognition (FER) has been considered significantly because of its wide range of applications, from lie detection to automatic counseling systems. The major part of our daily communication is consists of facial expression. 45% of daily communication between people is done by using language, and 55% is done by using facial expression [1]. FER has turned lots of attention to itself for a long time, and various methods have been used so far in order to detect a person's facial expression. Marketing research is one of its popular applications, where a customer's reaction is being observed without asking them to complete a survey about their purchase [2]. In this study we are going to use appearance-based methods to segment the mouth, then classify the facial expression based on the segmented region.

II. Literature Review

In this paper we are going to focus on mouth's shape or position, which will indicate facial expression. To perform feature extraction on the mouth or lips, the region must first be segmented. Lip segmentation is very challenging due to weak edges between lips and skin. In addition, other factors such as lighting, illumination or low resolution could also make lips segmentation difficult. Many methods such as Active Contour, Active Shape and Deformable Methods have been used so far in order to segment lips, but only work well under controlled circumstances [1] [3].

A new shape-constrain feature-based active contour (SCFAC) model, which shows a very good result achieving an F-score of 90.62% [4]. Using the analysis of colour in the image and a shape prior for the lips, the advantage of this method over others is that it can work well even under bad lighting and illumination [4]. As the study was focused on the segmentation method, the use of the segmented lips in FER was not considered.

Another solution used to segment the lips is the use of region of interest (ROI) detection using object detection and thresholds, which provided sufficient information to show the progression of fatigue in an individual [5]. The simplicity of the method enabled analysis of video data and the segmented lips was used for further analysis of fatigue in this case. This is similar to our objective to use the segmented region in FER of using the segmented region for classification. More complex and accurate methods can be used to better segment the data to provide more information to the classification step of the process.

III. Proposed Methodology

We propose a method that classifies facial expressions as either happy or sad using only the mouth. The algorithm was limited to features of the mouth because happiness and sadness can usually be evaluated by the shape of the person's mouth, or whether the individual is smiling. The first step in the proposed method is the identification of the region of interest, which is the area of the image around the mouth. This is performed by a Haar-Cascade objection detector. Next, the cropped area defined by the region of interest is passed to the segmentation method that produces a binary labelling of image. The desired foreground area should consist of the exclusively the mouth, which consists of a single topology containing the lips and visible teeth, and the background consisting of all other parts

of the image. The two segmentation methods considered in this report are a saliency map of the image and a level set algorithm using the Mumford-Shah functional. The segmented region is then used to train a CNN classifier to classify the image as either happy or sad. The purpose of the steps prior to the CNN classifier is to limit the information that the CNN uses to the image of the mouth, so that the model is trained only on relevant features. Segmentation methods will be evaluated using the Dice coefficient given by equation 1.

$$DSC = \frac{2|X| \cap |Y|}{|X| + |Y|} \tag{1}$$

The ground truth for segmented images was created manually for 10 images each for happy and sad expressions for a total of 20 images. The Dice coefficient is evaluated using the whole image, and not for the cropped image determined by the Harr-Cascade object detector, although the Dice coefficient does not change in either case.

A. Description of Dataset

In this study we use the Cohn-Kanade Au-Coded Expression Database for images to test and train our segmentation and classification method. There are 210 people participating in this study, who have an age range between 18 to 50 years and almost 2 out of three participants are female. A majority of the images are grey-scale with a small subset of images in colour. Images are taken from a frontal view with the face in roughly the same position for all participants. A series of images are taken starting with a neutral expression, and progresses to the targeted expression. Not all images in the sequence will be used in this study, as the scope of the study has been narrowed to happy and sad expressions to simplify the classification problem. The selected images are labelled as either happy and sad, and are manually segmented to create a ground truth for the purpose of evaluating segmentation algorithms. 200 images were chosen for each of the two expressions. The primary selection criterion was whether the image clearly depicted a happy or sad expression.

B. Haar-Cascade Object Detection

The Haar-Cascade is an ensemble of weak classifiers using the Adaboost learning algorithm developed by Paul Viola and Michael Jones [6]. Each single weak classifier evaluates a large number of simple features in all locations in the image. Because only a small minority of features are relevant, classifiers are trained using the Adaboost learning algorithm to identify the most relevant ones. All classifiers are combined into cascades, which consist of multiple stages made up of multiple classifiers of increasing complexity. An example must have a positive result for the first stage to be evaluated on the next set of classifiers. If an example has a negative result on any stage, the example is immediately evaluated as negative. This allows the cascade to evaluate general features that may indicate the desired object, and progressively use more specific features to determine if the object exists in the image. The purpose of the cascade is to improve the prediction time of the algorithm by only evaluating as many features as necessary to determine a negative result. We used a pre-trained Haar-Cascade to first identify the face in the image. The image was cropped to the identified face, then a second classifier was used to identify the mouth and its surrounding region. This outputted area was treated as the region of interest which was cropped and used in the following segmentation steps instead of the original uncropped image. In executing this procedure, only the region in the image around the mouth was used for segmentation to reduce error in the following steps, since unnecessary information in the image such as other facial features and background have been removed. Because only this region of the image will be used for segmentation, it reduces the number of other possible objects the segmentation algorithm will detect.

C. Segmentation by Saliency Map

Using the cropped images from the previous object detection step, one segmentation method that was tested was to create the saliency map of the image. Saliency detection describes a set of methods that highlight parts of the image that are likely to be features in the image based on the likely focus of a human observer [7]. The goal of this step is to aid feature extraction by the CNN algorithm for classification through providing only the relevant, segmented regions of the image. The saliency map is used so that likely features are outputted at a higher intensity, while the background is given a lower intensity. Then, a binary threshold is used to segment the features from the rest of the image. Fine grain static saliency is one method that attempts to mimic human vision by calculating on-center and off-center differences in the image to highlight important parts of the image [7]. Given a specific size of the kernel, regions of the image where the center has a large difference in intensity than the surrounding local area have a high on-center or off-center difference. Visual representations of the on-center and off-center differences kernels are shown in Fig. 1 [7]. Fine grain saliency accomplishes this saliency mapping at a higher resolution than previously implemented methods, and helps to exaggerate important parts of the image.

Spectral static saliency is another saliency method which uses the residual spectrum of the image to identify important features in the image [8]. This method determines the prior or expected information as the image's amplitude spectrum

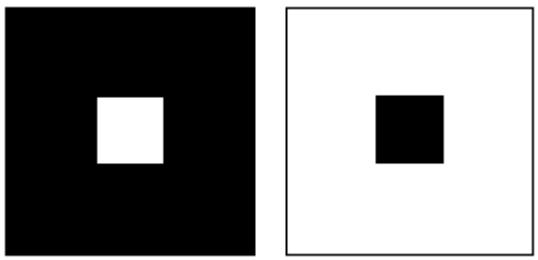

Fig. 1. On-Center and Off-Center Difference Kernels

filtered by a average filter in frequency domain, and is then subtracted from the original amplitude spectrum to determine the residual spectrum. This is then used to create the saliency map [8]. The resulting features and details arise in the saliency map from distinct parts of the image's amplitude spectrum that would be eliminated with an average filter. These parts of the amplitude spectrum indicate unique parts in the image that are not found anywhere else, which should be the desired features of the image. These are the two saliency methods that were considered and evaluated.

Using the highlighted features given a higher intensity from the saliency map, a binary threshold is then used to segment only the mouth from the image. Otsu's method is used to automatically determine a threshold that best separates the image histogram of the saliency map. All pixels greater than the threshold are labelled as the mouth, while the rest are considered as the background of the image. Both fine grained and spectral saliency were tested for their effectiveness in segmentation and classification.

D. Segmentation by Level Sets

Level sets are a common method of representing the contour of an object by non-parametric means. A level set identifies the contour in an image where the curve denoted phi is equal to zero. The curve is then evolved and minimized using an energy function as an active contour that defines the boundary of the segmented object. This curve evolution is based on equation 2. This segmentation method was chosen based on the results of [4] that used a shape prior in addition to the commonly used Mumford-Shah functional.

$$\phi_{i+1} = \phi_i + \Delta t \frac{\partial \phi}{\partial t} \tag{2}$$

The curve is iteratively evolved using this process, with the intended convergence of the contour at the boundary of the object defined by an energy function. After each iteration, the curve is redefined as the euclidean distance from the contour determined by the fast marching method (FMM) in the implemented algorithm. The curve evolution is related to the energy function as shown below.

$$\frac{\partial \phi}{\partial t} = -\frac{\partial E}{\partial \phi} \tag{3}$$

The energy function used for this segmentation is the Mumford-Shah functional with a shape prior. The Mumford-Shah functional uses four energy terms in the energy the function which describe the curvature of the contour, the length of the contour, and the inner and outer uniformities in intensity. The addition of the shape prior term directs the contour to be similar to the specified shape, which comes from prior knowledge. So the curve evolution is given by equation (4) [4], with the energy terms listed in the order they were described.

$$\frac{\partial \phi}{\partial t} = \delta(\phi)(v\kappa(\phi) - (I - c_1)^2 + (I - c_2)^2 - \mu - \tau C(\phi, \Phi_0)) \tag{4}$$

For simplification, instead of the equation used by [4], the energy term that utilizes the shape prior is given by the minimum distance from a given point on the contour defined by the level set to reference contour defined by the shape prior. The shape prior is an arbitrarily drawn approximation of the mouth. A shape prior representative of the dataset could not be created due to limitations of time, where many manual segmentations of the mouth are required to create the prior.

E. Classification

A convolutional neural network (CNN) was used to classify the segmented mouth images as either happy or sad. The primary advantage of CNN's over other classification methods for images is that it finds features of our dataset by itself rather than the designer defining the features manually. Additionally, spatial interactions between pixels are preserved due to the convolution operations performed by the network, compared to the method of using every individual pixel intensity as a feature. For this application, the working dataset is small with approximately 200 images. With a relatively small number of training examples, a simple model with a small number of parameters should be chosen. Therefore, we chose a CNN with Keras framework and sequential structure that consists of three convolutional layers, and a single fully connected layer. Hidden layers utilize a rectified linear activation function, and the computed loss is binary cross entropy.

The code used is a CNN code developed by Rajeev Ratan [9], which was originally used to classify dog from cat images, and was modified it for this application. The network was trained to label images as either happy or sad, instead of dog or cat in the original implementation. 25% of the images was used as the test set. The other 75% of the data were used to train the model, performing 25 epochs through the data during the training process. Data augmentation is a method that can help maximize the size of dataset by modifying images in the dataset. The Keras framework takes advantage of the ImageDataGenerator function to perform data augmentation, which performs a number of different geometric transformations such as translations, reflections, rotations, and scaling. Data augmentation was done as it was known that a dataset of 400 images is relatively small for deep learning purposes, so this was done to increase the size of the training set without requiring more data.

Multiple sets of images will be used to train the classifier. The first set of images will utilize the original uncropped images, so that the effectiveness of cropping the image can be observed by comparing the accuracy of this trained model to the others. The second set of images will be the cropped mouth images. For each saliency method, one model will be trained on the actual saliency map of the image, while the other will be trained on the segmented image determined by thresholding the saliency map. For two sets of images for each of the two saliency methods, four CNN classifiers will be trained. Lastly, one model will be trained on the cropped images segmented by the level set method. A total of seven CNN models will be trained using each of the seven different sets of images. For each data set, a second model will be trained using data augmentation. The purpose of training multiple

classifiers on different combinations of prior processing is to determine which methods benefit the classifier by improving its prediction accuracy, and which methods harm its accuracy.

IV. Results and Discussion

A. Haar-Cascade Object Detection

The Haar-Cascade object detection method was able to identify the region around the mouth in the image with an error of 7.0% out of the 402 images that were tested. Errors were defined as cropped images that did not include the mouth in the output image. Out of the errors encountered, 18 errors consisted of the Haar-Cascade identifying the region around the eye instead of the mouth as the region of interest. 10 errors was the failure to identify any area in the image as the mouth. Of the 20 images in the validation set to evaluate segmentation effectiveness, 5 incorrectly cropped the image. These images will be ignored in the mean calculation the Dice coefficient when evaluating the different segmentation methods. Examples of correctly cropped images are shown in Fig. 2 and 3, and images that identified the eye instead are shown below in Fig. 4.

We are limited in the options to correct these errors because a pre-trained algorithm is being used. One option is to retrain the Haar-Cascade using the specific dataset used in this project, so that it learns images that are relevant to this application. This option is very time consuming, as it would require labelled images of the mouth which is time consuming to create. Another option is to limit the search area for a mouth to the lower half of the identified face region. This would likely remove errors that identified the eyes instead of the mouth. However, instances where the Haar-Cascade fails to identify any region as the mouth will still occur using this solution.

Fig. 4. Haar-Cascade Incorrect Cropping for Happy Image

B. Saliency Map

1) Fine Grained Saliency: Fine grained static saliency was used to create a saliency map of each of the cropped images. Some examples of the resulting images are shown in Fig. 5 and 6. This saliency map appears to be effective in emphasizing features, as the edges around the mouth of the individual appear to be clearer than in the original images. When considering segmentation purposes, although the mouth is the brightest part of the image, the surrounding regions are gray. This can hinder segmentation when binary thresholding is applied to the image, as these regions may be included in the segmentation. Examples of the resulting thresholded images are shown in Fig. 7, 8, 9, and 10. The determined mean Dice coefficient using the validation set of images is 28.3%, which is a very poor result. Although fine grained saliency may appear to be effective in highlighting features of the image and making details of the image more visible, it performs badly when used with thresholding to segment features.

2) Spectral Saliency: Spectral saliency was also tested as an alternative to fine-grained saliency. Two examples of resulting images are shown in Fig. 11 and 12. The clarity in mouth features from spectral saliency vary between images. In Fig.

Fig. 2. Haar-Cascade Effective Cropping for Happy Image

Fig. 3. Haar-Cascade Effective Cropping for Sad Image

Fig. 5. Fine Grained Static Saliency for Happy Image

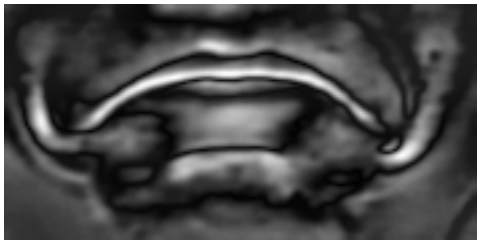

Fig. 6. Fine Grained Static Saliency for Sad Image

Fig. 7. Fine Grained Saliency Effective Segmentation for Happy Image

Fig. 8. Fine Grained Saliency Effective Segmentation for Sad Image

Fig. 11. Clear Result for Spectral Static Saliency for Happy Image

Fig. 12. Poor Result Spectral Static Saliency for Sad Image

11, the mouth appears to be much brighter than the rest of the background. For Fig. 12, the mouth is not visible at all, and the image as a whole appears very dark. This occurred in both happy and sad images. Like the previous saliency method, Otsu's method of thresholding was used to segment the image. Mathematical morphology was applied to the resulting binary mask as post processing, as it was observed that the edges were very jagged and the inside of the mouth was often missing. A circular mask of radius 11 was used for the dilation of the image, and a circular mask of radius 5 was applied for erosion. The resulting images are shown in Fig 13, 14, 15, and 16. Overall, spectral saliency works better than fine-grained saliency with thresholding to segment the mouth as indicated by the higher mean Dice coefficient of 47%. However, there were a multiple cases similar to Fig 16, where the output image was almost an entirely black image. This is caused by saliency maps similar to 12, where the features of the mouth appear very dark in the image.

C. Level Set Segmentation

The level set method described previously was executed for 200 iterations for each image. The contour was initialized as the rectangle 5 pixels from the border of the image. Parameter values were adjusted by trial and error based on the resulting segmented images which originate from the training set. They were initialized at the same parameter values in the method proposed by [4], and were further adjusted to match this

Fig. 9. Fine Grained Saliency Poor Segmentation for Sad Image

Fig. 10. Fine Grained Saliency Poor Segmentation for Sad Image

Fig. 13. Spectral Saliency Effective Segmentation for Happy Image

Fig. 14. Spectral Saliency Effective Segmentation for Sad Image

Fig. 15. Spectral Saliency Poor Segmentation for Sad Image

Fig. 16. Spectral Saliency Poor Segmentation for Sad Image

application by trial and error. The final values are shown in Table I. The shape prior used by the energy function is shown below in Fig. 17. Additionally, morphological closing using a circular mask of radius 3 was performed to smooth the edges of the contour as a post-processing step.

TABLE I
LEVEL SET PARAMETERS

Parameter Name	Parameter Description	Parameter Value
v	curvature of contour	0.1
μ	length of contour	500
λ_1	inside uniformity	1
λ_2	outside uniformity	1
τ	shape prior	0.05

Samples of the resulting segmented images with the original images are shown in Fig. 18, 19, 20 and 21. Fig. 18 and 19 show good results of the level set, where generally only the mouth is shown after applying the binary mask. However, there are significant portions of the tested dataset where the binary mask included regions that were not the mouth. Fig. 20 and 21 show segmentations that failed to segment the mouth, which were quite common when examining the output images. The mean Dice coefficient was found to be 24.7%. The low Dice coefficient indicates our algorithm performs extremely poorly in segmenting the mouth. Possible improvements that can be made to the level set algorithm is to use the same energy term for the shape prior as [4], which was not used for simplicity in the energy calculation. An additional energy term that is related to the magnitude of the image gradient at the contour can be added to direct the curve evolution to converge at the edges in the image. Further tuning of parameters may also be able to improve the result.

Fig. 17. Shape Prior for Level Set Segmentation

Fig. 18. Level Set Effective Segmentation for Happy Image

Fig. 19. Level Set Effective Segmentation for Sad Image

D. CNN Classifier

The CNN was trained using seven different conditions of the same dataset as listed previously. Data augmentation was performed for each of the sets, which doubled the number of conditions the CNN was trained on. The resulting loss and test

Fig. 20. Level Set Poor Segmentation for Sad Image

Fig. 21. Level Set Poor Segmentation for Sad Image

accuracy after 25 epochs of training for each model is shown in Table II.

TABLE II
LOSSES AND ACCURACIES FOR CNN CLASSIFIERS

CNN Model by Image Set	Unmodified Data		w/ Data Augmentation	
	Loss	*Accuracy*	*Loss*	*Accuracy*
Original Images	0.07	0.98	0.36	0.85
Cropped Images	0.20	0.95	0.09	0.97
Fine-Grained Original	0.22	0.98	<0.01	0.96
Fine-Grained Threshold	0.82	0.93	0.02	0.92
Spectral Original	1.03	0.91	0.93	0.84
Spectral Threshold	1.26	0.91	0.06	0.91
Level Set	0.27	0.92	0.13	0.90

While having better accuracy, models that did not use data augmentation were not reliable due to the relatively small dataset of 400 images. Evidence of this variation can be seen in graphs that plotted the training and test accuracy across epochs, as seen in Fig. 22 and 23 as one example.

The classifier trained using the original images had the highest accuracy when evaluated on the test images. This is likely because the classifier was able to use other features in the image that may indicate an individual's facial expression, such as the eye or other contours of the face that would be excluded in the cropped image. However with data augmentation, the test accuracy decreased by 10%. With a larger image, the additional images created by data augmentation had a greater variety than those created from a smaller image, so the CNN had more difficulty in finding relevant features across all images. For a consistent set of images like the Cohn-Kanade used in these experiments, additional processing of images prior to a CNN classifier does not benefit its accuracy very much, as it can already obtain an excellent accuracy of 98%. However, for a less consistent set of images, where the individual's face may be turned or rotated, the results of these tested methods such as object identification of the mouth and segmentation indicate a possibility to improve the test accuracy in these circumstances. This is indicated in the generally higher accuracies of models with data augmentation that use mouth images produced by the methods studied in this report. Only the model that used the spectral saliency map images had a lower accuracy. All other classifiers had an accuracy greater than %90.

Out of all pre-processing methods used to modify the original set of images, the CNN trained on the saliency map of the fine-grained saliency method resulted in the highest accuracy of 98% for models without data augmentation. With data augmentation, the model trained on the cropped images had the highest accuracy. Models trained with unsegmented data generally had a higher accuracy than those trained with segmented images. Part of this reason is likely due to the poor performance of the segmentation methods, as indicated by the mean low Dice coefficients that were reported in this paper. To confirm whether segmentation is actually a useful method to help train a CNN classifier, a model can be trained using manually segmented data. However, there was insufficient time to manually segment enough images to train the CNN.

V. CONCLUSION AND FUTURE WORK

A convolutional neural network was able to classify facial expressions as happy or sad with high accuracy using the image region around the mouth which was found by the use of a Haar-Classifier. One model that uses fine grained saliency maps of the region surrounding the mouth reached a test prediction accuracy of 98%. Segmented images of the mouth was also used to train a CNN to perform this classification. The segmentation methods used to generate the training images were thresholding of fine grained and spectral saliency maps, and a level set that uses the Mumford-Shah functional with a shape prior. However, all methods performed poorly when evaluated against 20 manually segmented images by the Dice coefficient. The highest mean Dice coefficient was achieved by thresholding the spectral saliency image, with a value of 47%. When these segmented images were used to train a CNN classifier, the accuracy of the trained classify was lower

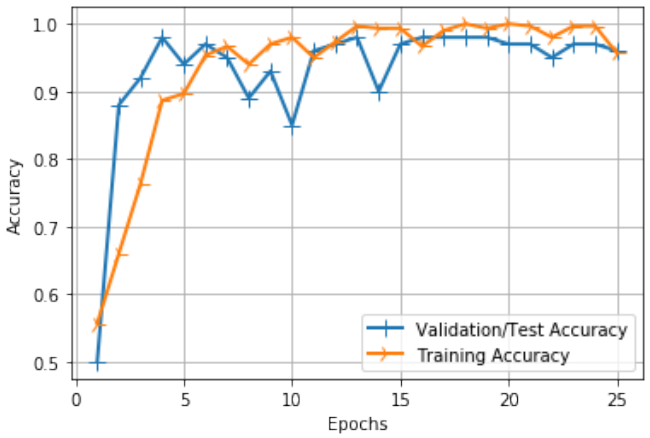

Fig. 22. Test and Training Accuracy for Model Using Cropped Images over 25 Epochs

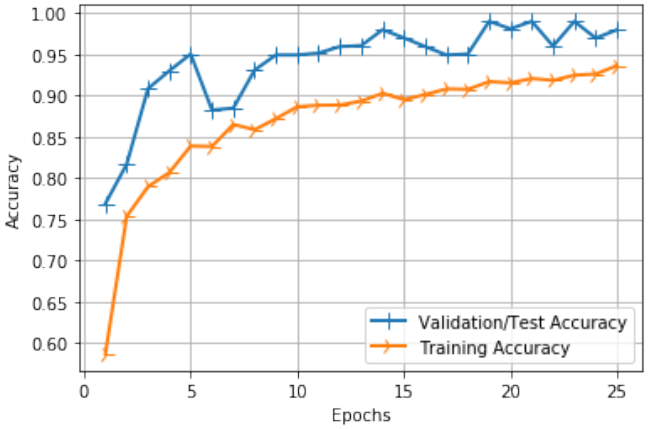

Fig. 23. Test and Training Accuracy for Model Using Cropped Images with Data Augmentation over 25 Epochs

than models trained on unsegmented images. This neither confirms or denies that using segmented images can improve classification, as the lower accuracy can likely be attributed to the poor performance of these segmentation algorithms, and its failure to properly segment the mouth.

Future work can be done to improve the segmentation process before the CNN classifier. Because of insufficient time, an expansive set of manually segmented images could not be created. A large set of manually segmented data can help improve the current methods of segmentation. With enough training data, the Haar-Cascade can be trained on the actual dataset instead of using a pre-trained classifier. This can help reduce the failure rate and incorrect identifications of the region of interest. A more representative shape prior can also be used to improve the level set segmentation using manually segmented images instead of using an arbitrary shape as the reference. The equation for the shape prior term that was not implemented for simplicity in the level set algorithm can also be used to improve the segmentation and obtain a result closer to [4]. The set of manually segmented images can also be used to train a CNN classifier to determine the use of segmented images improves the accuracy of a classifier. Evidence could not be determined for this due to the poor performance of the tested segmentation methods.

ACKNOWLEDGMENT

Thanks to Dr. Eran Ukwatta for his support and guidance through this project and in teaching fundamentals of image processing in ENGG*6090.

REFERENCES

[1] C. Bouvier, P. Y. Coulon, and X. Maldague, "Unsupervised lips segmentation based on ROI optimisation and parametric model," *Proceedings - International Conference on Image Processing, ICIP*, vol. 4, 2007.

[2] N. Samadiani, G. Huang, B. Cai, W. Luo, C. H. Chi, Y. Xiang, and J. He, "A review on automatic facial expression recognition systems assisted by multimodal sensor data," *Sensors (Switzerland)*, vol. 19, no. 8, pp. 1–27, 2019.

[3] M. M. Hasan and M. F. Hossain, "Facial features detection in color images based on skin color segmentation," *2014 International Conference on Informatics, Electronics and Vision, ICIEV 2014*, pp. 1–5, 2014.

[4] T. H. N. Le and M. Savvides, "A novel Shape Constrained Feature-based Active Contour model for lips/mouth segmentation in the wild," *Pattern Recognition*, vol. 54, pp. 23–33, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.patcog.2015.11.009

[5] S. R. Khanal, A. Fonseca, A. Marques, J. Barroso, and V. Filipe, "Physical exercise intensity monitoring through eye-blink and mouth's shape analysis," *TISHW 2018 - 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing, Proceedings*, 2018.

[6] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE Comput. Soc, 2001, pp. I–511–I–518. [Online]. Available: http://ieeexplore.ieee.org/document/990517/

[7] S. Montabone and A. Soto, "Human detection using a mobile platform and novel features derived from a visual saliency mechanism," *Image and Vision Computing*, vol. 28, no. 3, pp. 391–402, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.imavis.2009.06.006

[8] ——, "Human detection using a mobile platform and novel features derived from a visual saliency mechanism," *Image and Vision Computing*, vol. 28, no. 3, pp. 391–402, 2010.

[9] R. Rajeev, "Data Augmentation - Cats vs. Dogs," 2019.