

Detection of Visual Positive Sentiment using PCNN

Samar H. Ahmed¹, Emad Nabil², Amr A. Badr³

Computer Science Department
Faculty of Computers and Information
Cairo University, Egypt

Abstract—Many people all over the world use online social networks to express their feeling and sharing their experience, and the easiest way from their perspective is using images and videos to do so. This paper shows the utilization of two techniques (Viola et al algorithm and Pulse coupled Neural Network) in visual sentiment analysis using a hand-labeled dataset. The proposed system, which uses the PCNN with NN classifier, achieves 96% right classification, whereas Viola algorithm achieves 94% for the same dataset.

Keywords—Visual sentiment analysis; Pulse Coupled Neural Network (PCNN); viola et al. algorithm

I. INTRODUCTION

Nowadays online Social networks sites have a great role in people lives for communicating and to exchange any information with each other, including their opinions, feelings and any perspective regarding several topics in our life. We can find all this huge knowledge embedded in different aspects, such as tags and comments on the social networks as well as microblogging sites. In the last decade analyzing and understanding emotion and sentiment from user's content including (text, video, and images) has become a great role in behavior study. Such information can be used in a wide range of applications such as business intelligence applications, political, and the stock market. Sentiment analysis from visual content has attracted many researchers since the sentiment that can be extracted from the visual content such as (video or images) can highly explain the sentiment compared with text sentiment. The analysis of such information can strongly be useful for real live applications such as a rating for places using this visual content, predicting accidents based on humans reactions captured by street cameras, also it can be used to measure people's satisfaction in streets and places to provide such information to the ministry of happiness.

The objective of this research paper is to experiment two different techniques for extracting sentiments from images. The first technique is inspired from Viola et al. work [1], in which they use the Haar-like features to detect faces. An extension to this technique is added to allow the algorithm the sentiment within these images. Whereas in the second technique the visual sentiment extraction is done using the concept of Pulse Coupled Neural Network (PCNN) [2]. These two techniques will be compared using the same dataset and the results will be discussed.

The rest of the paper is organized as follows. Section 2 describes related work. Section 3 introduces our proposed method. Section 4 provides results and discussion, followed by conclusion in Section 5.

II. RELATED WORK

The authors in paper [3] have investigated the connection between images of natural scenery and the human understanding of an image semantics utilizing the Ortony, Clore & Collins (OCC) emotion model [4]. A solid classifier was built by incorporating the AdaBoost calculation in addition to the Back Propagation neural system (BP neural system) strategy, which brought about the automatic emotion classification of natural scenery. As a result of their proposed solution demonstrated that the AdaBoost-BP neural network algorithm achieved mean recall and precision rates of 91.5% and 86.7%, individually, for natural scenery semantic classification, which demonstrate an expansion of 3.5% and 4.2%, compared with contrasted and the mean recall and precision rates of the BP neural system classification algorithm (88.0% and 82.5%, respectively).

The authors in paper [5] have proposed another CNN architecture that completely utilizes joint text-level and image-level portrayal to perform mixed media sentiment analysis. In light of thought of the correlative impact of the two portrayals as sentiment features, the proposed strategy takes the benefit of the inner connection amongst text and image in image tweets and uses it to accomplish better performance results in sentiment prediction. They also proved that their solution accomplishes better results than different algorithms like SVM, Logistic regression over two different tweeter datasets.

The authors in paper [6] used the deep learning techniques to classify emotion based on quite large dataset from flicker and tweeter, trying to classify it based of the main five categories (*Love, Happiness, Sadness, Violence, and Fear*). The results of their method demonstrated that deep learning provides promising outcomes with performance compared with methods that utilizing high-quality features on emotion classification process.

The authors in [7] and [8] proposed two different methods to utilize current visual content including attributes as features for image sentiment analysis. But the main obstacle of their proposed methods was at defining the mid-level attribute point in the training phase as this approach requires huge domain knowledge of linguistic as well as human intervention to fine-tune the results.

The authors in [9] proposed a novel sentiment analysis structure in the light of convolutional neural network for visual sentiment analysis prediction. They demonstrated that the image representations from the CNN trained on a huge scale dataset could be effectively transmitted for sentiment analysis.

The authors in [10] used the approach proposed in [9] that utilizing small scale datasets, which are totally different in nature from the pre-trained image dataset, resulting better performance by 4.5% than the proposed approach in [9].

The authors in [11] realized the importance of extracting a sentiment from both textual and visual content for understanding the human sentiment and get an accurate sentiment labeling. In this work, they used CNN for extracting the sentiment out of textual content and DNN for extracting the sentiment for the visual content. The evaluation for the author’s framework was done on a dataset, which was collected from a famous Chinese social network called “Sina Weibo”.

The authors in [12] demonstrated that the parameters of CNN that are trained on large-scale dataset (e.g., ILSVRC dataset) can be transmitted to object detection and scene classification conducting preferable performance over classic handcrafted representations.

III. PROPOSED METHODS

This section presents in much more details two techniques that are utilized for extracting sentiments from visual content. The first technique, that will be called *technique (1)* in the rest of this paper, was originally suggested by Viola et al. to detect faces from a set of images contains faces and non-faces images. This technique is refined to extract the sentiment from the visual content.

Whereas the second technique, that will be called *technique (2)* in this paper, utilized the concept of Pulse-coupled neural network for image smoothing, image segmentation and feature extraction. This technique is extended to perform the task of extracting a sentiment from visual content. The following sub sections will describe these two techniques:

A. Technique (1): Application of Haar-Like Features

1) *Haar-like features* [13]: Haar-Like features are features that are used to detect digital image objects using the concept of Haar basis functions. A subset the features that are used in this paper is shown in “Fig. 1”. As shown in this figure, the first feature is the two rectangles located horizontally or vertically. This feature is defined as the difference between the summations of pixels of these two same size rectangles. The second feature is Three-rectangle feature where its value is obtained by subtracting sum of pixels under two outside rectangle and from sum of pixels under center rectangle. The third feature is a four- rectangle feature, its value is computed by subtracting the diagonal pairs of rectangles.

Viola system consists of three components namely: *Preprocessing, Feature Extractor and cascaded classifier*. (see “Fig. 2”). The following section describes each in much more details.

2) *Preprocessor component*: The objective of this component simply is to prepare the images to be processed by the feature extractor component that is working based on

Haar-like features presented above. The preprocessor will operate according to the following steps:

a) Convert images to greyscale image since Haar-like features cannot operate in colored images.

b) Scale down images to 24*24 pixels.

c) Transform each image to its corresponding integral image (ii). The purpose of using the concept of integral image is to save computation time

d) Compute the values of Haar-like features using the concept of integral image (ii) [1] and create a table named Summed Area Table. In this table, at any point (x, y) in the original image $i(x,y)$ there is a corresponding value $ii(x,y)$. This value itself is the sum of all the pixels values above, to the left and of course including the original pixel value of (x, y) itself. The computation formula is as follows:

$$ii(x,y)=$$

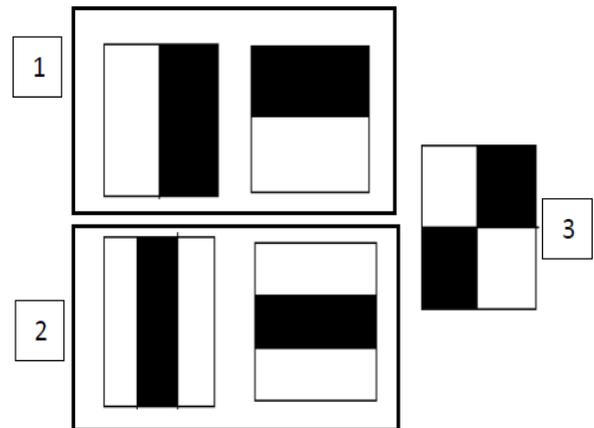


Fig. 1. Example of Haar Like Features.

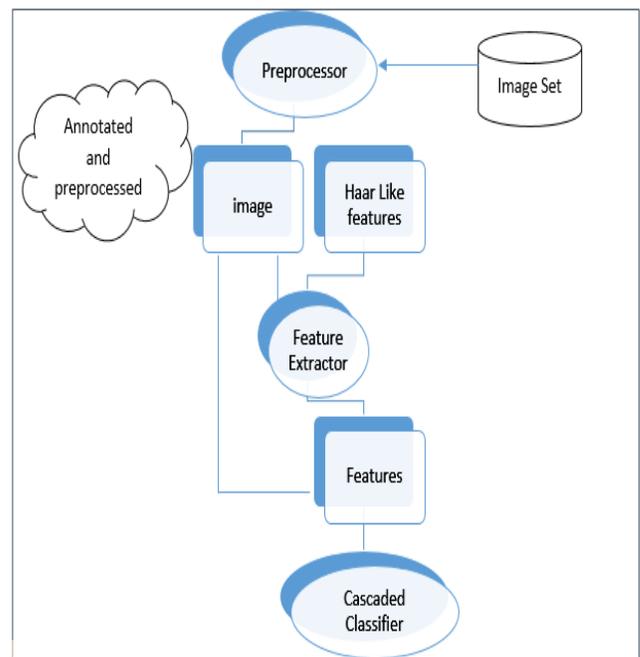


Fig. 2. Viola et.al Algorithm Structure.

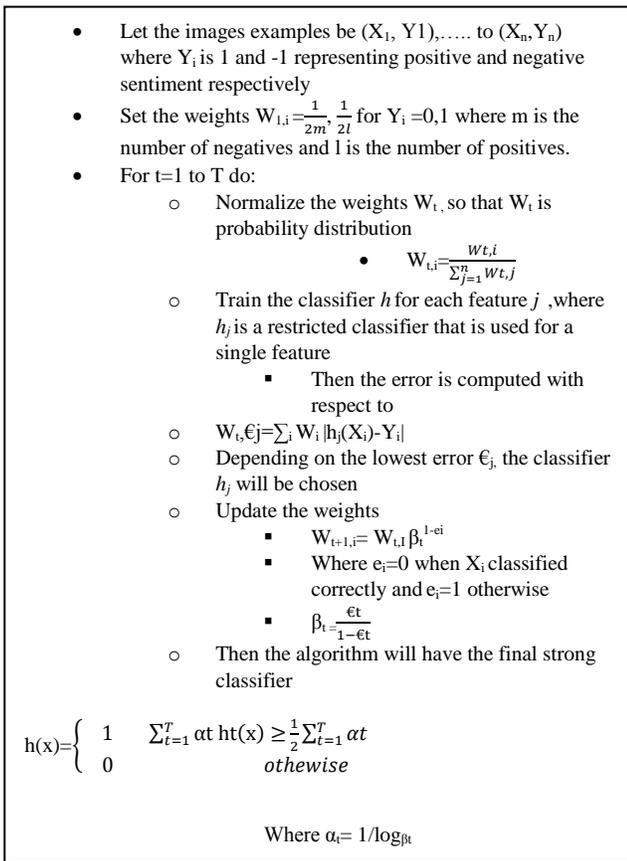


Fig. 3. Adapted AdaBoost Algorithm (Adapted From [1]).

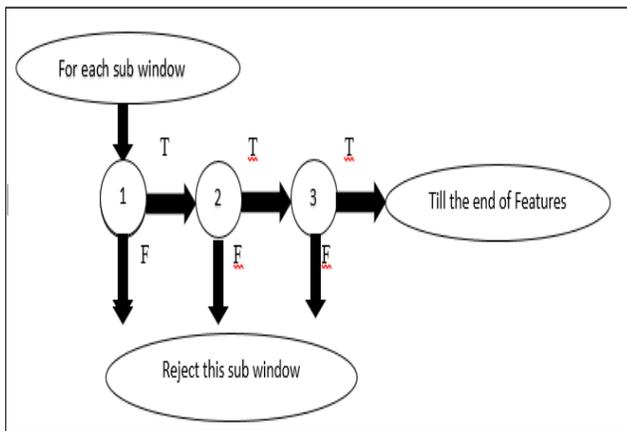


Fig. 4. Schematic Depiction of the Detection Cascade [1].

Feature Extractor Component: The objective of this component is to extract set of features that can be found in a very large set of the Haar-like features obtained from the previous component. To build a strong classifier that best describes a face with sentiment, a variant of AdaBoost algorithm [1] is used both to select a small set of features and train the classifier (see “Fig. 3”).

Cascaded Classifier component: The objective of this component is to increase detection performance while radically minimizing computation time by generating a

decision tree called *cascade* (see “Fig. 4”). For each image sub window a series of classifiers are applied with cascading. The positive result from the first weak classifier triggers the second weak classifier and so on till the end of all weak classifiers (features) that are existed in the system’s strong classifier generated by the previous component. If a negative result was reached at any point an immediate rejection of this sub window is done.

B. Technique (2): Application of Pulse Coupled Neural Network

This technique offers visual sentiment extraction using both Pulse Coupled Neural Network (PCNN) algorithm and NN classifier (see “Fig. 5”). The reasons for choosing PCNN is its ability to best describe the image features by generating an image signature.

For each image, the PCNN algorithm runs for 50 iterations resulting 50 deviated binary images (BI) based on the formulas parameters initiated with zero while the theta notation initiated with 0.1. Each image will have a sequence of signatures Sig_{ij} where j is the image number and i is the BI number.

An image signature is a result of counting the number of 1’s in each binary image. As a result, a data file of the generated sequences will be created that is annotated manually by assigning 1 for positive images and -1 otherwise. This generated file will be uses in NN classification process.

As shown in “Fig. 6” PCNN is working as follows [2]:

PCNN neuron: Feeding and Linking are two primary compartments. Each compartment has an entangled connection with neighboring neurons with M and W weights separately. Each holds its past state changed by a rot factor. Once the Feeding compartment gets the info boost, S ; each compartment is calculated by the following formula (1, 2):

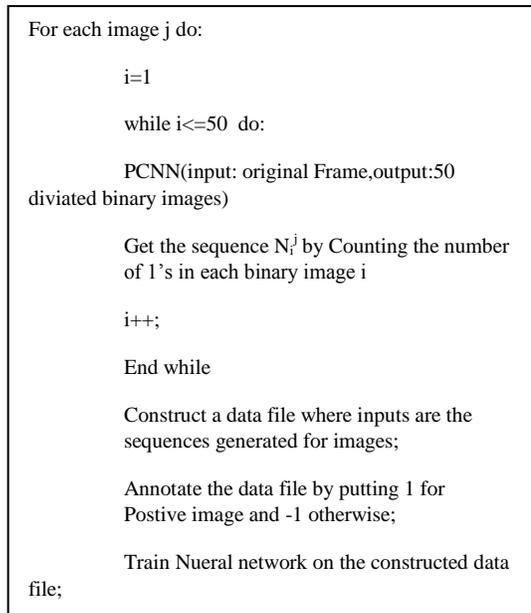


Fig. 5. Technique (2) Application of PCNN Algorithm.

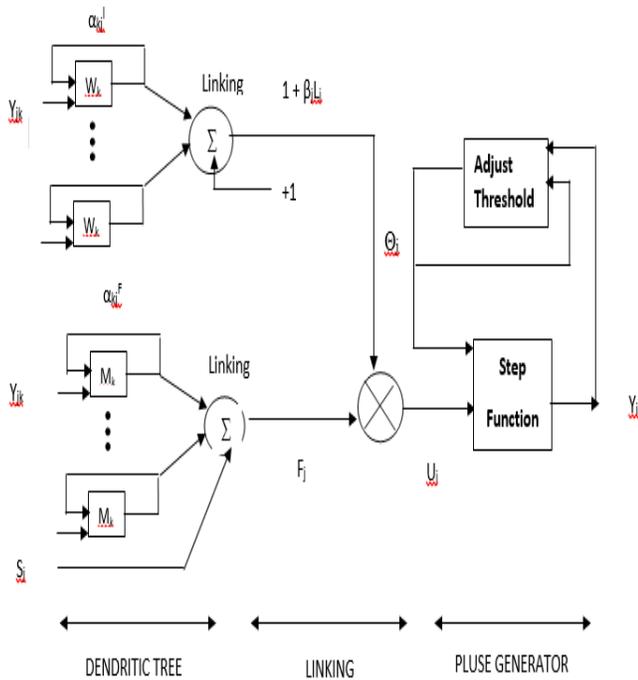


Fig. 6. Semantic Representation of PCNN Processing Element [2].

$$F_{ij}[n] = e^{\alpha F \delta n} F_{ij}[n-1] + S_{ij} + VF \sum_{kl} M_{ijkl} Y_{kl}[n-1], \quad (1)$$

$$L_{ij}[n] = e^{\alpha L \delta n} L_{ij}[n-1] + VL \sum_{kl} W_{ijkl} Y_{kl}[n-1], \quad (2)$$

Where:

- F_{ij} : is the Feeding compartment of (i,j) neuron.
- L_{ij} : is the Linking compartment of (i,j) neuron.
- Y_{kl} : is the output neuron from the last iteration [n-1].
- $e^{\alpha F \delta n}$,and $e^{\alpha L \delta n}$: are exponent terms for both compartments that are used to rot the previous (n-1) state through the time (n).
- VF and VL : are constant values that are used for normalization. So if the weights of M and W have any changes, these values are used to prevent saturation by scaling the resultant correlation.

Each state of feeding and linking compartments respectively are banded together so it can create an internal state of the neuron which called U . the term β is the linking strength that is used to control this combination, this internal state is calculated by the following formula (3):

$$U_{ij}[n] = F_{ij}[n] (1 + \beta L_{ij}[n]), \quad (3)$$

The output Y is produced by comparing the internal state of the neuron by a threshold Θ (see formula (4)):

$$Y_{ij}[n] = \begin{cases} 1 & \text{if } U_{ij}[n] > \Theta_{ij}[n-1] \\ 0 & \text{Otherwise} \end{cases} \quad (4)$$

This threshold is controlled by formula (5):

$$\Theta_{ij}[n] = e^{\alpha \Theta \delta n} \Theta_{ij}[n-1] + V\Theta Y_{ij}[n], \quad (5)$$

Where:

$V\Theta$: is a large constant that is generally greater than the average value of U_{ij} .

IV. RESULTS AND DISCUSSION

To train the predictor, a balanced dataset consisting of 200 images, downloaded from online social networks, is divided into two equally hand labeled groups: smiley faces representing positive sentiment and non-smiley faces representing negative sentiment. The positive group is further divided into two subgroups: the first subgroup has 50 images with strong positive sentiment and the second subgroup has 50 images with nature positive sentiment. On the other hand the second group that represents the negative sentiment further divided into two subgroups: the first subgroup consists of 50 strong negative images and the second subgroup consists of 50 nature negative images. A sample of the visual dataset is shown in "Fig. 7".

The dataset is divided equally for training part and testing part.

The experiment includes the following steps:

1. Download images from online social networks.
2. Classify and manually annotate images into two groups (Positive and negative as explained above).
3. Train both technique (1) and technique (2) for the set on the annotated images.
4. Run both technique (1) and technique (2) for the test set of the images measuring the percentage of correct classification.
5. Table 1 summarizes the results obtain from running the experiments on different number of images in range (100 to 200).



Fig. 7. Sample the Dataset used in Experiments.

TABLE I. COMPARISON OF ACCURACY OBTAINED FROM THE TWO TECHNIQUES

	Number of images used per experiment				
	100	120	150	170	200
Technique (1):App of Haar-like features	90%	91.6%	92%	91.7%	94%
Technique (2):App of PCNN	92%	93.3%	94.6%	95.2%	96%

V. CONCLUSION

This paper experiments two techniques for extracting the sentiment from face images. The first technique uses Haar-Like Features, where the second technique which presents a new architecture using the concept of pulse coupled neural network (PCNN) with NN classifier. The experiments were done on a balanced dataset which contained 200 face images downloaded from online social networks. The second technique shows a better and consistent accuracy (96%) as it can work with any image with different sizes and colors, while it suffers from high computations and therefore requires more processing time in addition to its inability to improve its accuracy unless its parameters are optimized using an optimization algorithm such as genetic algorithm. On the other hand, technique (1) shows less accuracy and less processing time. Also, technique (1) suffers from constraints on images as it must be converted to grayscale mode and rescaled to 24 X 24 pixel.

REFERENCES

- [1] P.Viola. And M.Jones,(2001), "Rapid object detection using a boosted cascade of simple features. In Proceedings", IEEE Conference on Computer Vision and Pattern Recognition, ISSN: 1063-6919.
- [2] T.Lindblad, J.M. Kinser, (2005), "Image processing using pulse-coupled neural networks". New York, NY: Springer, Book ISBN, 978-3-642-36877, pp. 146-151.
- [3] J.Cao, J. Chen, and H.Li,(2014), "An AdaBoost-Backpropagation Neural Network for Automated Image Sentiment Classification", The Scientific World Journal, vol 2014, Article ID 364649, 9 pages.
- [4] A.Ortony, G. L. Clore, and A. Collins, (1998), "The Cognitive Structure of Emotions", Cambridge University Press, Cambridge, UK.
- [5] G. Cai, B. Xia, (2015), "Convolutional Neural Networks for Multimedia Sentiment Analysis". In: Li J., Ji H., Zhao D., Feng Y., "Natural Language Processing and Chinese Computing", Lecture Notes in Computer Science, vol 9362, Springer, Cham.
- [6] V.Gajarla and A.Gupta, (2015), "Emotion Detection and Sentiment Analysis of Images", Georgia Institute of Technology.
- [7] J. Yuan, S. McDonough, Q. You, and J. Luo,(2013), "Sentribute: Image sentiment analysis from a mid-level perspective," in Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, vol 13, pp. 10:1-10:8.
- [8] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang,(2013), "Large-scale visual sentiment ontology and detectors using adjective-noun pairs," in Proceedings of the 21st ACM International Conference on Multimedia, vol 13, pp. 223-232.
- [9] C. Xu, S. Cetintas, K. Lee, and L. Li, (2014), "Visual sentiment prediction with deep convolutional neural networks," arXiv preprint arXiv, CoRR, vol. abs/1411.5731.
- [10] S. Jindal, S. Singh, (2015), "Image sentiment analysis using deep convolutional neural networks with domain-specific fine-tuning", International Conference on Information Processing (ICIP), Vol 736, pp. 447-451, International Conference on. IEEE.
- [11] Y. Yuhai, L. Hongfei, M. Jiana, Z. Zhehuan.,(2016) "Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks". Algorithms, issue 9 vol 2.
- [12] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell,(2013), "Decaf: A deep convolutional activation feature for generic visual recognition," vol. abs/1310.1531, pp 647-655.
- [13] S.Sriram, B.Illuri, (2016) "Real Time Smile Detection using Haar Classifiers on SoC", vol 104 - No 10, International Journal of Computer Applications, pp 30-34.
- [14] J.Derrac, S.García D,Molina , F. Herrera,(2011),. "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms". Swarm and Evolutionary Computation, Issue 1, vol 1, pp 3-18.

Although technique (2) shows that it has more processing time since PCNN requires high computations. On the other hand it shows a better accuracy than technique (1) by 2% for 200 images (see "Fig. 8"). In addition PCNN shows that it has consistent results over different number of images unlike technique (1).

It is also noticing that technique (2) is invariant with image rotations and resizing as the image signature will remain the same every time the given image gets rotated or resized. Unlike technique (1), technique (2) has an advantage as it can work efficiently with any image size so there is no need to rescale images as technique (1) does. Technique (2) can also work with both colored and grayscale images not only with grayscale images as technique (1) does.

To check if there is a significant difference between the two techniques, the Wilcoxon's rank-sum test was performed [14] for PCNN against Haar-Like Features. The output of the test is a p-value. This p-value determines the significance level of the two algorithms. If the P-value is less than 0.05, there will be a significant difference. After running the Wilcoxon's rank-sum test, the P-value equals 0.042, which means that the performance of the application PCNN really outperforms the application Haar-Like Features technique.

However, Technique (2) is not able to accomplish a higher accuracy that already accomplished since the image signature will remain the same every time unless the PCNN formula parameters are fine-tuned.

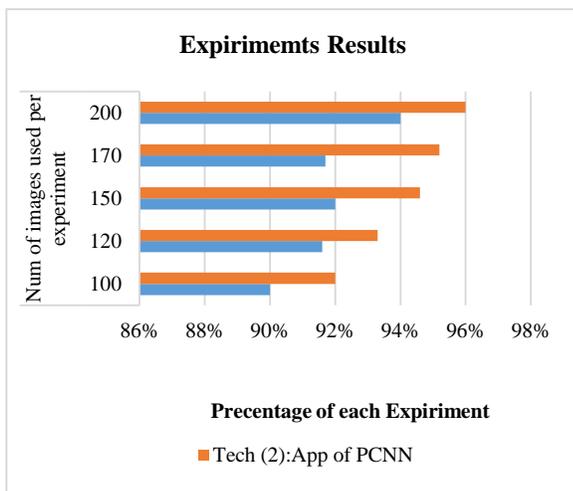


Fig. 8. Comparison between Two Techniques in a Graphical Format.