International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V69I11P102

Volume 69 Issue 11, 6-10, November 2021

Review Article

Implementation of Python Packages for Image Recognition

Snehal Shah[1], Kishan PS[2], Jitendra Jaiswal[3]

[1,2] *Computer Science with specialization in Data Science Engineering(Hons)*
[3]*Faculty of Engineering and Technology - Jain University, Kanakapura Taluk, Ramanagara District, Karnataka, India.*

Received Date: 24 September 2021
Revised Date: 26 October 2021
Accepted Date: 08 November 2021

Abstract – *In today's society, data plays a significant role from time to time. Here we have taken the real-time image as well as the video for recognition of objects. Using CNN (convolutional neural network) to recognize images, machine learning and deep learning play a crucial role in object recognition. We have used YOLO for object detection, where the images and framework are divided into grids. The OpenCV package allows real-time recognition of images and videos. NumPy package is used for calculation purposes and to check the confidence level and even FPS of the image or video*

Keywords - *YOLO, OpenCV, NumPy, CNN, DNN.*

I. INTRODUCTION

What is Image Recognition? The process of identification and detection of an object while showing an image, or a video, or a live video is known as Image Recognition. Image recognition like an object, people, targeted objects, animals, etc., can be done, which will be helpful in real-time application. In this generation, image recognition is very important. By using Image recognition, we can achieve a lot of things; for example, we can control traffic violations, Crime cases can be reduced, finding criminals will become easy, etc.

Machine learning and deep learning can be used to tackle image identification. Identifying and extracting significant elements from photos and applying them as a machine learning model is what image recognition entails. Convolutional neural networks (CNNs) may be used in deep learning to detect such features in fresh photos. Analyzing, visualizing, and developing algorithms within a technical computing environment is a powerful approach for image recognition.

II. LITERATURE SURVEY

Deep learning (also known as deep structured learning or hierarchical learning) is a type of machine learning that uses artificial neural networks. Learning can be supervised, semi-supervised, or unsupervised. [1][2][3] Computer vision, speech recognition, natural language processing, audio recognition, social network filtering, machine translation, bioinformatics, drug design, medical image analysis, material inspection, and board game programming have all used deep learning architectures such as deep neural networks, deep belief networks, recurrent neural networks, and convolutional neural networks.[4][5][6] There are multiple layers between the input layer and the output layer of a deep neural network (DNN).[7][2]DNN determines the correct mathematical manipulation formula to turn any relationship into an output, whether it's a linear or non-linear relationship. The network traverses the layers, assessing the likelihood of each output in the example of a dog picture; a DNN trained to detect dog breeds may examine the image and determine the chance that the dog belongs to a specific breed. After reviewing the findings, the user may choose which probabilities should be presented (those that are greater than a specific threshold, for example) and return the recommended label. In DNNs, each layer corresponds to a mathematical transformation, and therefore the term "deep" refers to the number of layers. Image recognition has become superhuman because to deep learning technologies, which give results that are more accurate than those produced by people.[8]Vehicles with deep learning training can now comprehend 360° camera images. Another example is FDNA, which analyses instances of the human deformity using a big database of genetic disorders. Convolutional neural networks (CNNs or ConvNets) are a form of deep neural network used to assess visual pictures. Because of their shared-weights architecture and translation invariance qualities, they're also known as shift invariant or space invariant artificial neural networks (SIANN).[9][10][11] They have applications in image and video recognition, recommender systems,[12] image classification, medical image analysis, and natural language processing.[13] A system to recognize hand-written ZIP Code numbers[14] involved convolutions in which the kernel had been laboriously hand-designed.[15] It was shown by K. S. Oh and K. Jung that GPUs could greatly accelerate standard neural networks. They achieved 20 times faster performance on GPUs than on

CPUs.[16][17]. Another research from 2005 emphasized the need for GPGPU for machine learning.[18]K. Chellapilla et al. published the first GPU implementation of a CNN in 2006. They were four times faster to implement than a CPU-based version.[19]Following that, GPUs were employed for other types of neural networks (not just CNNs), particularly unsupervised neural networks.[20][21][22][23] When used to facial recognition, CNNs resulted in a significant reduction in mistake rate.[24]Another study found that "5,600 still photos of more than 10 participants" had a recognition rate of 97.6%.[25]

III. MATERIALS AND METHODS

A. Machine learning

Machine learning is an artificial intelligence technique that allows computers to learn and improve without having to be explicitly programmed. The construction of computer programs that can access data and learn on their own is referred to as machine learning.

Computational statistics, which focuses on making predictions with computers, is closely connected to machine learning. Machine learning benefits from mathematical optimization research because it provides tools, theory, and application fields.

Data mining is a field that deals with machine learning. Machine learning is also known as predictive analysis when used to solve business challenges.

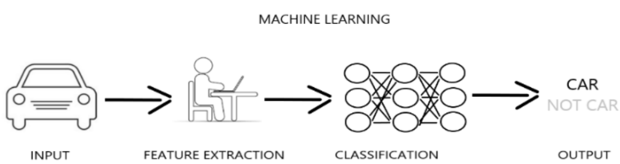

Fig. 1 Machine learning

B. Deep Learning

Deep learning, also known as deep structured learning or hierarchical learning, is a subset of artificial neural network-based machine learning algorithms. There are three types of learning: supervised, semi-supervised, and unsupervised.

Deep learning is a category of machine learning techniques that employs numerous layers to extract higher-level features from raw data. Lower layers in image processing, for example, may recognize edges, whereas higher layers may identify human-relevant notions like numerals, letters, or faces.

Artificial neural networks, such as convolutional neural networks, are often used in contemporary deep learning models.

Deep learning techniques reduce feature engineering in supervised learning tasks by translating data into compact, intermediate representations analogous to principal components and constructing layered structures that eliminate representation redundancy.

Unsupervised learning difficulties can benefit from deep learning methodologies. Because unlabeled data is more numerous than labeled data, this is a considerable benefit.

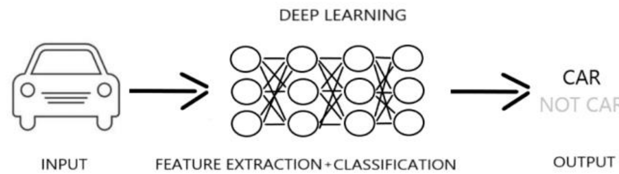

Fig. 2 Deep learning

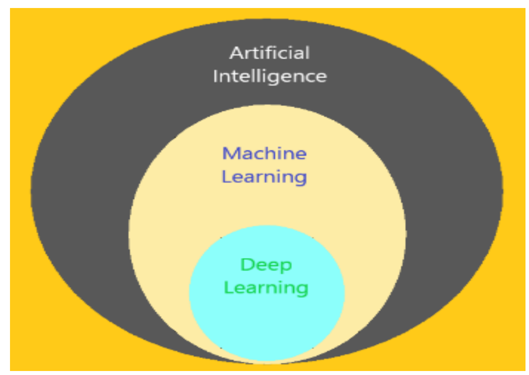

Fig. 3 Structure of AI, DL, ML

C. Neural networks

In more contemporary terms, a neural network is a network or circuit of neurons or a synthetic neural network made up of artificial neurons or nodes. These artificial networks may be used for predictive modeling, adaptive control, and other applications that require a dataset to train. Self-learning may happen in networks, which can draw inferences from a complicated and apparently unconnected set of data.

The idea of neural networks has been used to better understand how neurons in the brain work as well as to provide the groundwork for artificial intelligence research.

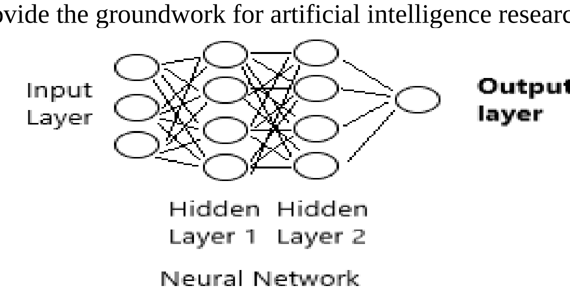

Fig. 4 Neural Network

D. Image Recognition

An image recognition algorithm (also known as an image classifier) takes an image (or a patch of an image) as input and outputs the image's contents. To put it another way, the result is a class label (for example, "cat," "dog," "table," and so on). How can an image recognition algorithm figure out what's in a picture? To discover the distinctions between various classes, you must train the algorithm. If you want to find cats in photos, you'll need to train an image recognition algorithm using thousands of shots of cats and thousands of images of backdrops that don't include cats. Needless to say, this algorithm can only comprehend the objects/classes that it has been taught.

To keep things simple, we'll simply look at two-class (binary) classifiers in this piece. You would assume that this is a fairly restrictive assumption. However, many popular object detectors (such as the face detector and pedestrian detector) use a binary classifier. A face detector, for example, has an image classifier that determines if a region of an image is a face or a background.

E. CNN

The abbreviation for convolutional neural network (CNN) is the convolutional neural network. CNNs are built up of neurons with learnable weights and biases, similar to neural networks. Before reacting with an output, each neuron processes a weighted sum of many inputs using an activation function. As the name implies, neural networks are machine learning system that is modeled after the organization of the brain. It is made up of a network of neurons, which are learning units. The cornerstone of automated recognition is the learning of how to transform input signals (such as an image of a cat) into corresponding output signals (such as the label "cat"). Take, for example, computer-assisted image identification. Take, for example, computer-assisted image identification. To detect whether or not a picture contains a cat, an activation function is utilized. The label "cat" will be activated if the image resembles prior cat photographs that the neurons have seen.

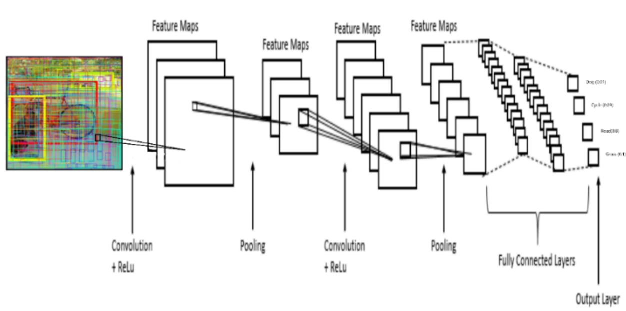

Fig. 5 CNN structure

F. YOLO

YOLO is You Only Look Once. Use for object detection. It is a completely different approach. It looks at the image in a clever way just for once. It will work by dividing the images or frameworks into grids. YOLO uses deep learning and convolutional neural networks to detect objects (CNN). As a result, YOLO is one of the most rapid means of detection. This graphic, for example, demonstrates how YOLO works.

It will divide this image into s × s grids.

Each cell is connected with the five bounding boxes.

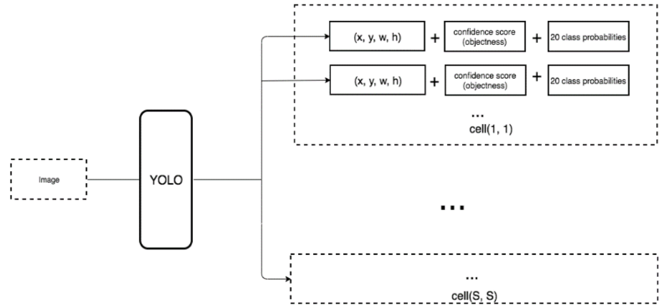

Fig. 6 YOLO architecture

Fig. 7 Normal Image

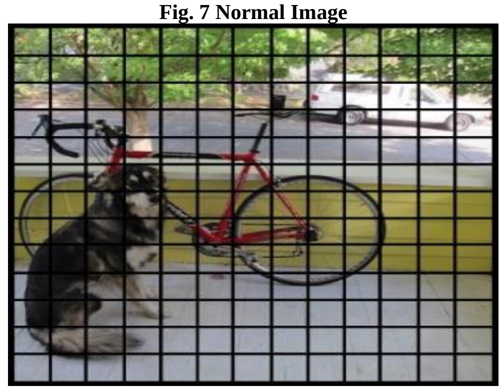

Fig. 8 Grid View

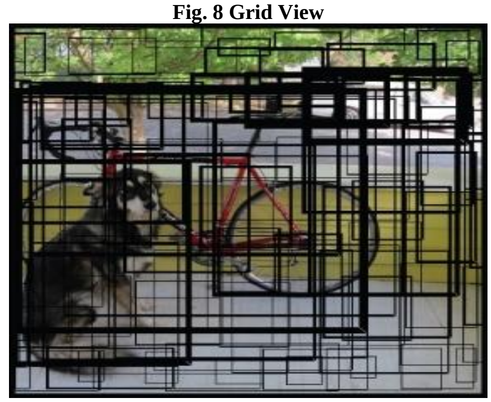

Fig. 9 Detected Object in Image

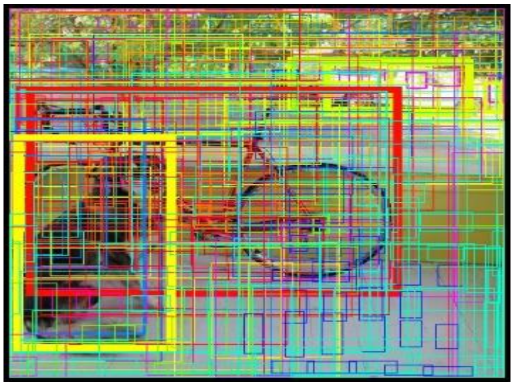

Fig. 10 Detected image in different bounding box

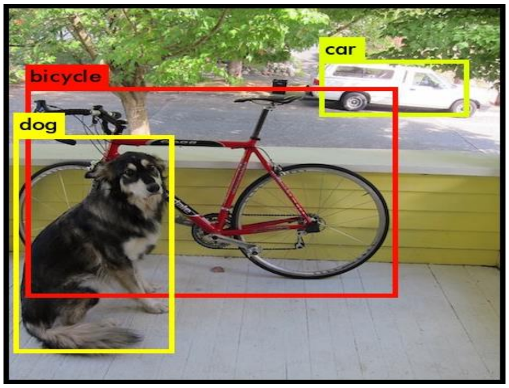

Fig. 11 Removing of the extra bounding box and also detection of the perfect image

G. OpenCV

The acronym OpenCV refers to an open-source computer vision library. It's mostly used in real-time computer vision applications. TensorFlow, PyTorch, and Caffe are some of the deep learning frameworks supported by OpenCV. OpenCV is a cross-platform framework for real-time computer vision application development. It primarily focuses on image processing, video recording, and analysis, with features like face and object recognition thrown in for good measure. Even when using OpenCV, NumPy interprets picture data as ndarray, therefore remembering the processing in NumPy is helpful (ndarray).

H. NumPY

Numpy stands for Numerical Python. General-purpose is an array processing package. Multidimensional array object with high performance and tools for dealing with it. Python is used to do basic scientific computations. It's used for multi-dimensional array processing that isn't related to image processing. By the operation of ndarray, acquisition and rewriting of pixel values, trimming by the slice, concatenating can be done. Those who are used to NumPy can do a lot of things without using libraries such as OpenCV.

Image = 2-D numerical array

Numpy: basic array manipulation

IV. RESULT AND DISCUSSION

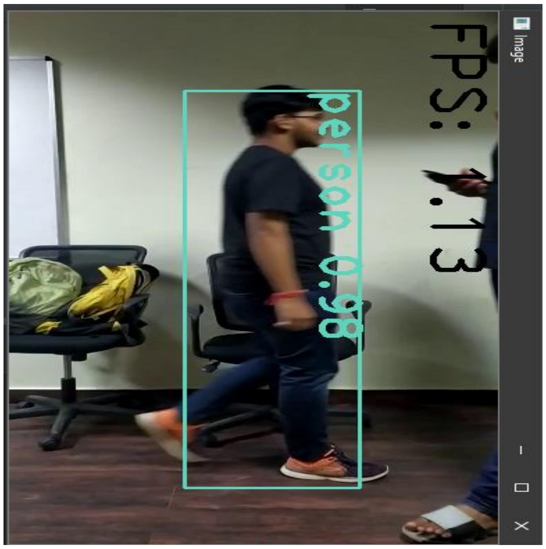

Fig. 12 Live Screenshot of video output – 1

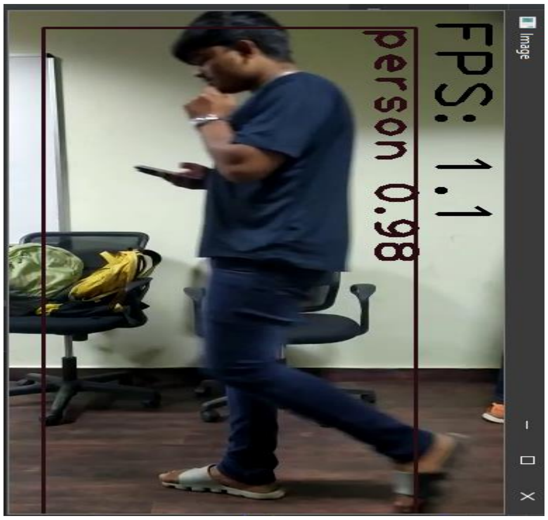

Fig. 13 Live Screenshot of video output – 2

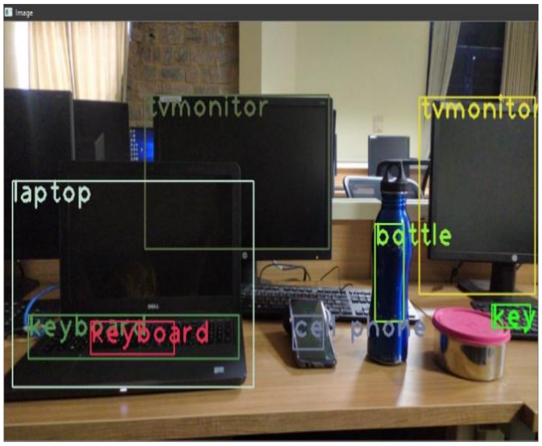

Fig. 14 Live Screenshot of image output

V. CONCLUSION

Working with YOLO and OpenCV packages that we have observed the effectiveness and importance of accuracy. Accuracy places a very important role in image and video recognition and plays the main role in the performance of object detection. There are several benefits to incorporating picture and video recognition in this fast-rising population and technology. Here, we have worked on improving accuracy from 90% to 98%, and it helps in detecting the object faster and more effectively than previous performance.

VI. FUTURE SCOPE

It can be used in a traffic violation. Like if a rider is not following the traffic rules and regulations and violating them, we can rise challan to that person very fast by saving time and work. It will initially assess whether the cyclist is wearing a helmet using image recognition. If the rider is not wearing a helmet, the camera will capture the bike registered number plate and generate the challan to the registered number plate. We can find the criminal by giving the image of the criminal to the database, and when the camera catches the real-time image of the criminal, it will verify with the database the image; if the image matches, then the location of the place will be sent to the cyber-crime. From this, the police will be able to capture the criminal easily. It will help in scanning the face faster than normal speed. Face recognition, security, and surveillance, as well as visual geolocation, object recognition, gesture recognition, code recognition, industrial automation, and image analysis in medical and driver assistance, are all important corporate applications of image recognition. As a result of these applications, several locations are experiencing greater growth.

REFERENCES

[1] Bengio, Y.; Courville, A.; Vincent, P. Representation Learning: A Review and New Perspectives, IEEE Transactions on Pattern Analysis and Machine Intelligence. 35(8) (2013)1798–1828.

[2] Schmidhuber, J., Deep Learning in Neural Networks: An Overview, Neural Networks. 61 (2015) 85–117.

[3] Bengio, Yoshua; LeCun, Yann; Hinton, Geoffrey.,Deep Learning, Nature. 521 (7553) (2015) 436–444.

[4] Ciresan, D.; Meier, U.; Schmidhuber, J., Multi-column deep neural networks for image classification, 2012 IEEE Conference on Computer Vision and Pattern Recognition., (2012) 3642–3649.

[5] Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffry., ImageNet Classification with Deep Convolutional Neural Networks(PDF), NIPS ,Neural Information Processing Systems, Lake Tahoe, Nevada, (2012).

[6] Google's AlphaGo AI wins three-match series against the world's best Go player, TechCrunch, (2017).

[7] Bengio, Yoshua., Learning Deep Architectures for AI(PDF), Foundations and Trends in Machine Learning, 2 (1) (2009) 1–127. CiteSeerX 10.1.1.701.9550. doi:10.1561/2200000006. Archived from the original(PDF) on 2016-03-04. Retrieved 2015-09-03.

[8] Cireşan, Dan; Meier, Ueli; Masci, Jonathan; Schmidhuber, Jürgen (August 2012).,Multi-column deep neural network for traffic sign classification, Neural Networks. Selected Papers from IJCNN, 32 (2011) 333–338

[9] Nvidia Demos a Car Computer Trained with Deep Learning (2015-01-06), David Talbot, MIT Technology Review, (2015).

[10] Zhang, Wei., Shift-invariant pattern recognition neural network and its optical architecture, Proceedings of Annual Conference of the Japan Society of Applied Physics,(1988).

[11] Zhang, Wei., Parallel distributed processing model with local space-invariant interconnections and its optical architecture, Applied Optics, 29(32) (1990) 4790–7. Bibcode:1990ApOpt..29.4790Z. doi:10.1364/AO.29.004790. PMID 20577468.

[12] van den Oord, Aaron; Dieleman, Sander; Schrauwen, Benjamin (2013-01-01). Burges, C. J. C.; Bottou, L.; Welling, M.; Ghahramani, Z.; Weinberger, K. Q. (eds.). Deep content-based music recommendation(PDF). Curran Associates, Inc., 2643–2651.

[13] Collobert, Ronan; Weston, Jason (2008-01-01). A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. Proceedings of the 25th International Conference on Machine Learning. ICML '08. New York, NY, USA: ACM, (2008) 160–167. doi:10.1145/1390156.1390177. ISBN 978-1-60558-205-4.

[14] Denker, J S, Gardner, W R., Graf, H. P, Henderson, D, Howard, R E, Hubbard, W, Jackel, L D, BaIrd, H S, and Guyon ,Neural network recognizer for hand-written zip code digits, AT&T Bell Laboratories, (1989).

[15] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, Backpropagation Applied to Handwritten Zip Code Recognition; AT&T Bell Laboratories.

[16] Oh, KS; Jung, K., GPU implementation of neural networks, Pattern Recognition, 37(6) (2004) 1311–1314. doi:10.1016/j.patcog.2004.01.013.

[17] Schmidhuber, Jürgen., Deep Learning, Scholarpedia. 10(11) (2015) 1527–54. CiteSeerX 10.1.1.76.1541. doi:10.1162/neco.2006.18.7.1527. PMID 16764513.

[18] Dave Steinkraus; Patrice Simard; Ian Buck., Using GPUs for Machine Learning Algorithms, 12th International Conference on Document Analysis and Recognition (ICDAR 2005) (2005) 1115–1119.

[19] Kumar Chellapilla; Sid Puri; Patrice Simard., High-Performance Convolutional Neural Networks for Document Processing , In Lorette, Guy (ed.). Tenth International Workshop on Frontiers in Handwriting Recognition, Suvisoft, (2006).

[20] Hinton, GE; Osindero, S; Teh, YW., A fast learning algorithm for deep belief nets, Neural Computation. 18(7): 1527–54. CiteSeerX 10.1.1.76.1541.doi:10.1162/neco.2006.18.7.1527. PMID 16764513, (2006).

[21] Bengio, Yoshua; Lamblin, Pascal; Popovici, Dan; Larochelle, Hugo., Greedy Layer-Wise Training of Deep Networks(PDF). Advances in Neural Information Processing Systems, (2007)153–160.

[22] Ranzato, MarcAurelio; Poultney, Christopher; Chopra, Sumit; LeCun, Yann (PDF). Advances in Neural Information Processing Systems, (2007).

[23] Surendiran, R. and Alagarsamy, K., PCA based geometric modeling for automatic face detection. Int. J. Comput. Sci. Inform. Technol, 1(4) (2010) 221-225.

[24] Raina, R; Madhavan, A; Ng, Andrew Large-scale deep unsupervised learning using graphics processors., (PDF). ICML, (2009) 873–880.

[25] Lawrence, Steve; C. Lee Giles; Ah Chung Tsoi; Andrew D. Back., Face Recognition: A Convolutional Neural Network Approach., IEEE Transactions on Neural Networks. 8(1) (1997) 98–113. CiteSeerX 10.1.1.92.5813. doi:10.1109/72.554195.

[26] Matusugu, Masakazu; Katsuhiko Mori; Yusuke Mitari; Yuji Kaneda., Subject independent facial expression recognition with robust face detection using a convolutional neural network.,(PDF). Neural, (2003).