**Honey Bees classification using Deep learning**

**Abstract:** This project includes the implementation of deep learning-based architectures for the classification of 4 styles of honeybees classification of bees was extremely needed to grasp the characteristics of them which may be accustomed verify and notice the health of the Bee hive's, honey bees square measure the foremost necessary Pollinators in our system. By this, we can boost the method of pollination which can end in high-quality food merchandise for North American country. standard CNN-based architectures were evaluated supported classification accuracy and prediction speed. supported the results an easy nonetheless effective Convolutional Neural Network (CNN) was projected with some high accuracy and a few tight prediction speed.

**1.Introduction.**

Every third bite of food depends on pollination by bees. the foremost necessary issue that bees do is fertilize. pollenation is required for plants to breed, and then several plants rely upon bees or different insects as pollinators. At a similar time, this past winter Apis mellifera hive losses have exceeded an hour in some states. however, will we tend to address this issue? however, will we tend to higher perceive our bees? will CNN create a bearing on our problem? can it ready to improve the pollination process? And most significantly, however, will we tend to save them before it's too late? whereas several indications of hive strength and health square measure visible on the within of the hive, frequent check-ups on the hive square measure long and riotous to the bees' progress and hive normally. By work the bees that leave the hive, we can gain an additional complete understanding of the hive itself. as an example, associate unhealthy hive infected with varroa mites can have bees with malformed wings or mites on their backs. These characteristics will be ascertained while not gap the hive. to guard against thief bees, we tend to might track the quantitative relation of pollen-carrying bees vs those while not. an outsized flow of bees while not spore is also a sign of thief bees. This information set aims to produce basic visual data to coach machine learning models to classify bees in these classes, paving the means for additional intelligent hive observance or cultivation normally. we tend to begin by finding out the dataset at hand. information preprocessing was done to form the information able to be used. First, information was increased to equalize category imbalance, and standardization was applied thereto. Models were trained and tested on this preprocessed information. Throughout this project, our main focus was to extend quality step by step and puzzle out what works the simplest for this dataset.

**2.Methodology.**

When the computer came into existence in the 1960s it aimed to mimic human behavior and its way of thinking, vision systems, and many more things and to ask them what do they think and see, this can be said to be automating the process of image analysis.

Just like animals see the world differently, behave differently from humans to external forces, same way computers has their way of seeing the world they differentiate each image or vision by the number of pixels, try to draw the border between the objects by considering various factors like shades of colors, spatial relationships between the objects.

As computer vision started to get evolved there came into existence many algorithms which can solve and improve the vision of computers at a very good scale, they started to be programmed and got better and better as we feed them more amount of data.

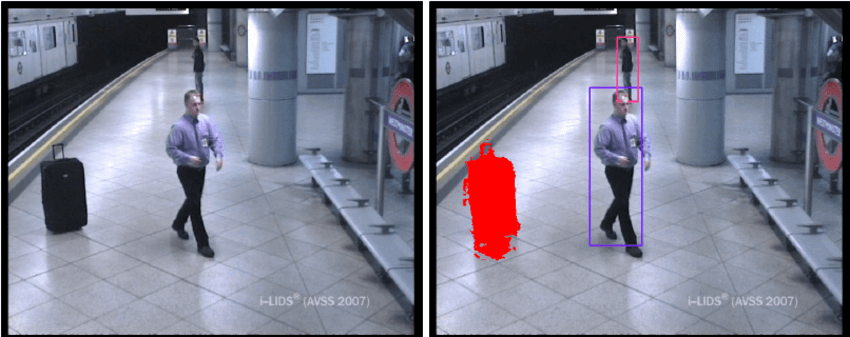


Fig 1.a) CCTV cameras detecting objects

Coming to the decade of 2010 we have seen enormous growth in the field of deep learning and computer vision. With the evolution of high computing hardware and deeper algorithms, we can train supercomputers to perform some extremely difficult tasks just in few minutes. Which can lead to improve our business solutions and can be highly profitable for big organizations.

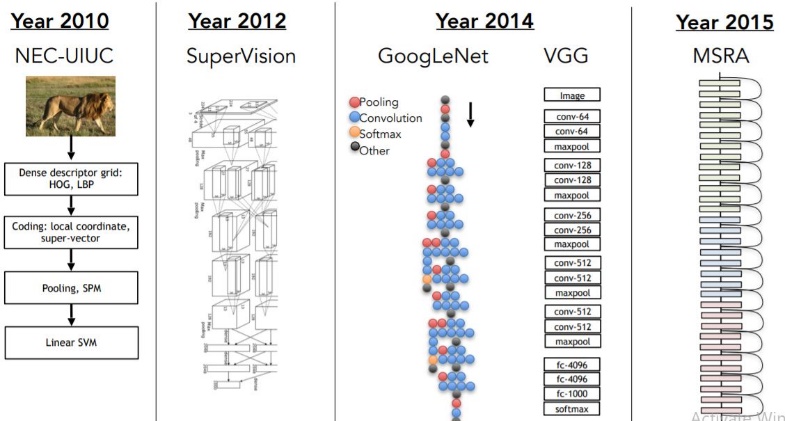


Fig 1.b) Evolution of Computer Vision

Nowadays we will extract 3-dimensional pictures with the assistance of our algorithms that wasn't doable in earlier days of laptop vision. Studies within the earlier 1970s fashioned the bottom of today's laptop vision algorithms including advanced arithmetic and convolutions which might notice edges, labeling of lines, sharping of pictures, adjusting brightness, contrast, representation of the combination of pictures, optical flow, the motion of the image, etc.

Now we are going to discuss concerning some state of art algorithms like Imagenet. Imagenet consisted of fifteen million pictures that were high resolution and belonged to twenty-two thousand classes. the pictures were collected from the net and tagged by human labelers mistreatment Amazon's Mechanical Turki crowd-sourcing tool. In 2010 a contest Image web giant-scale visual recognition challenge(ILSVRC).

In ILSVRC there area unit around one.2 million coaching pictures,50000 validation pictures, and one,20,000 testing pictures. On ImageNet, it's customary to report 2 error rates: top-1 and top-5, wherever the top-5 error rate is that the fraction of taking a look at pictures that the proper label isn't among the 5 labels thought of most probable by the model. Imagenet consists of varied resolution pictures, while we tend to need the same resolution of an image to feed to a neural network therefore we'd like to downsample the resolution to 256X256.

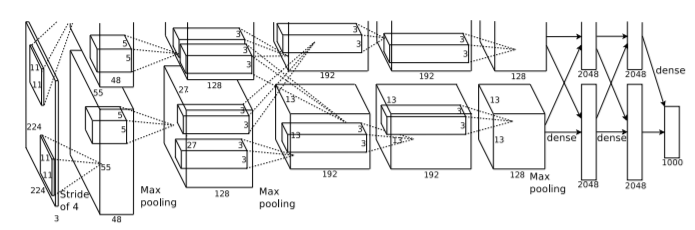


Fig 1.c) Architecture of Imagenet

Another state of art algorithmic program is Alexnet that is incredibly a lot larger as compared to previous algorithms. It has sixty million parameters with a half-dozen,50,000 neurons and it takes 5 to 6 days to coach Alexnet on higher computational GPU's like GTX 580, but this was back in 2012. Now we've got more advanced CNN's which might be dead quicker even on larger datasets.

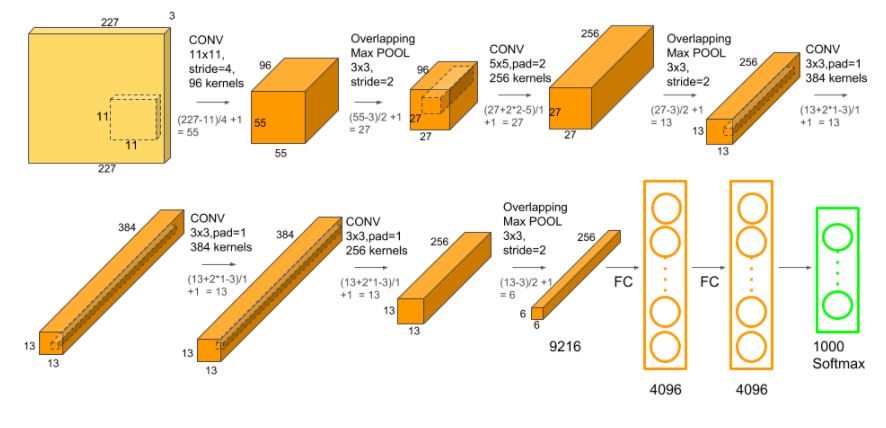


Fig 1.d) Design of Alexnet

Multiple convolutional filters are employed in the CNN to extract attention-grabbing options from our input image. In single convolution this id est multiplication of 2 functions there is several filters gift that extracts useful options from our dataset. The dimension and height of the kernel are sometimes identical and also the depth is that the same because of the variety of channels. The first convolutional layer is followed by a GHB pooling layer, the third, fourth, and fifth convolutional layers are connected directly. The fifth convolutional layer is followed by the Associate in Nursing Overlapping GHB Pooling layer, the output of which matches into a series of 2 connected layers. The last layer or connected layer feeds into a softmax activation perform for the classification of pictures, Relu non-linearity activation perform is applied to any or all the hidden layers that are followed by social control before feeding it to the pooling layer to scale back non-linearity.

The next decade saw some advanced and rigorous studies within the field of laptop vision that LED to some advanced mathematical analysis. This includes the analysis within the field on shading, texture, focus, contrast, contour, and lots of alternative options of a specific image.

The convolutional neural network is only hooked to the operation of convolution. The convolution operation uses a little size weight of matrix that ar captive through the method layer.

A pooling layer is applied to that for feature extraction. It will cut back the spatial size of a picture. It is primarily used once the convolutional layer. This is then fed to a completely connected layer where numerous weights are initialized once coaching convolutional networks, there are some difficulties that special coaching techniques are developed. for instance, since the degree of preparation, a model is set by each its power and also the quantity of coaching it receives, providing a convolutional network with a lot of coaching examples will cut back preparation. as a result of these networks sometimes train all offered information, one approach is to either generate new information from scratch, if attainable, or augment existing information to make new information.

Recent work has seen the revival of feature-based strategies, employed in conjunction with machine learning techniques and sophisticated improvement frameworks. The advancement of Deep Learning techniques has brought more life to the sector of laptop vision. The accuracy of deep learning algorithms on many benchmark laptop vision information sets for tasks starting from classification, segmentation, and optical flow has surpassed previous strategies.

The application of laptop vision is large in every and every field in today's world. Nowadays huge organizations are mistreatment this technology at an oversized scale. Recent work of Amazon led to the development of a method known as Virtual cause figurer wherever someone will try garments just about while not carrying it this has been attainable with the assistance of advanced CNN called GAN's that uses the idea of generators and decoders. In that, it'll observe the cause of humans and check the key points of the idea of image segmentation with it. This application of computer vision has brought huge advancement in the field of business in the retail sector.

Also, it's brought laurels within the field of automobile, according to World Health Organization over 1.2 million individuals dies thanks to road accidents which may be avoided with the assistance of laptop vison. The conception of self-driving cars is currently a reality with corporations like Tesla, Waymo area unit operating for the most part in the developments of those technologies. the corporate claims to use deep networks for prediction, planning, mapping, and simulation to coach the vehicles to maneuver through completely different things like construction sites, fall to emergency vehicles, create space for cars that area unit parking, and stop for crossing pedestrians.

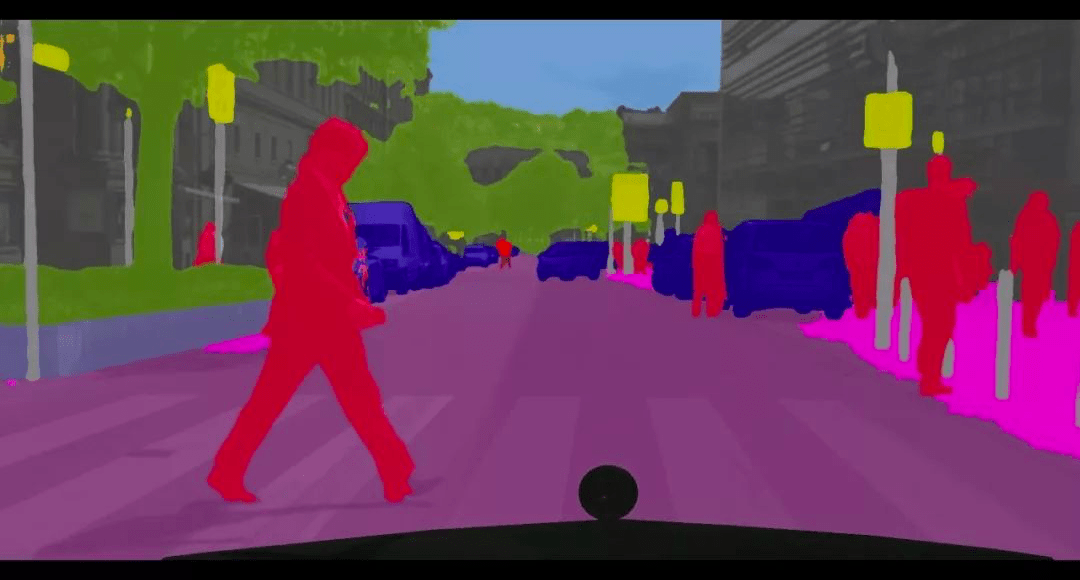


Fig 1.e) Image segmentation used in self-driving cars

The applications of computer vision have also been helpful in the field of the health care sector we can detect the cells and tissues. We can easily detect diseases by a study of cells and make a model with that to prevent the harmful effect.

Also, it has proven helpful in the field of agriculture to improve the quality of crops, our project deals with such application of computer vision on bee classification to improve the process of pollination which will help in better quality food items. These applications claim to offer farmers the opportunity to conduct precision farming, to raise production at a lower cost. Many advancements and research are still going in the field of Computer vision for the betterment of humanity.

**3.Simulations**

**3.1 Dataset Information**

The Dataset used here is Annotated HoneyBee Image Dataset from Kaggle which consisted of 4 categories of Bee images of dimension (72x72x3),3 signifying the RGB channel of colored images. The dataset consisted of 5173 images in which 4143 was used for training the model,500 for validation, and 500 for a test. Below are some sample images from our dataset.

Fig.3.1 a) Sample Images from Dataset

**3.2 Data Preprocessing**

In preprocessing we have done data augmentation. Images from low-frequency classes were picked up and random rotation and brightness variation was performed. Eventually, data was normalized for all classes. Then, standardization was performed on the data to normalize its mean and make it unit variance. Also, we need to convert the images into an array so that our deep learning algorithms can be applied and further can be fed to our ANN for forwarding and backward propagation. During training, the machine adjusts its internal parameters to project each feature tensor close to its target.

After training, the machine can be used to predict the target for previously unseen feature tensors. What this study focuses on is the requirement that feature tensors must be of the same size. In other words, the same number of features must be present for each sample. Hence we need to resize every image present in our dataset



array ([[[130., 165., 142.],

[130., 165., 142.],

[130., 165., 142.

[ 56., 82., 59.],

[ 56., 82., 59.],

[ 56., 82., 59.]],

[[130., 165., 142.],)

Fig 3.2 a) The above image is converted into an array to feed to a neural network

**3.3 Model Implementation**

In this section, we will discuss the CNN model implemented on our dataset for classification in detail.

Here we can also see if pixels have any sequential relation which can be exploited to boost the performance

**3.3.1 Proposed Convolutional Neural Network**

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets can learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

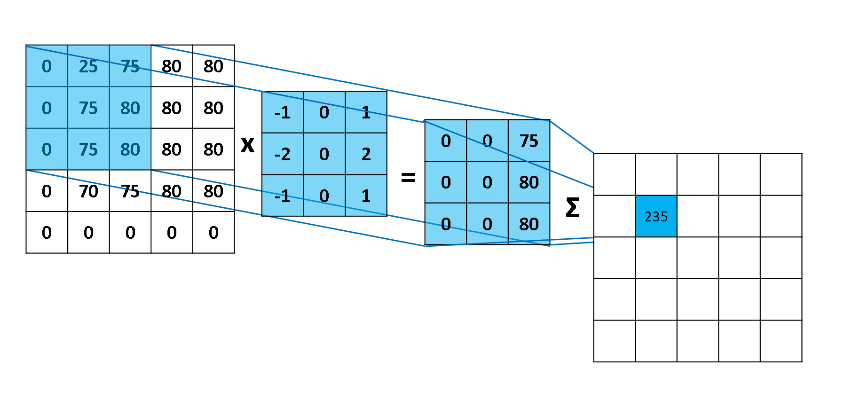


Fig 3.3.1.a) Convolution Operation in CNN

In Convolutional Neural Networks, Filters observe spacial patterns like edges in a picture by detective work the changes in intensity values of the image.

In terms of a picture, a high-frequency image is that the one wherever the intensity of the pixels changes by an oversized quantity, whereas a low-frequency image is that the one where the intensity is sort of uniform. Usually, a picture has high and low-frequency elements. The high-frequency elements correspond to the perimeters of AN object as a result of at the perimeters the speed of modification of intensity of element values is high. High pass filters are wont to enhance the high-frequency components of a picture.

Padding could be a term relevant to convolutional neural networks because it refers to the number of pixels supplemental to a picture once it's being processed by the kernel of a CNN.

An activation perform could be a vital feature of a man-made neural network, they decide whether or not the vegetative cell ought to be activated or not. In artificial neural networks, the activation perform defines the output of that node given AN input or set of inputs.

There are 2 varieties of Pooling are Max Pooling and Average Pooling. Pooling returns the utmost price from the portion of the image lined by the Kernel. On the opposite, Average returns the common of all the values from the portion of the image lined by the Kernel.

We enforced a convolution neural spec that has 2 stacks created from 3 convolution layers followed by a max-pooling layer once every convolutional layer and 2 connected layers. we tend to used batch normalization once each convolution layer to avoid exploding and vanishing of gradients whereas back-propagating across totally different layers ReLU activation perform was used as a nonlinearity once every convolution layer that set negative input to zero units. The convolution layers have sixteen, 32,64 convolution filters with kernel size = a pair of and learning rate=0.0001. Its output is given to a completely connected hidden layer consisting of five hundred neurons. the ultimate output layer has four neurons with softmax as activation perform to calculate the chance admire four categories of our dataset.

**4.Result's**

The test accuracy achieved here is 96.23% for our model with 20 epochs

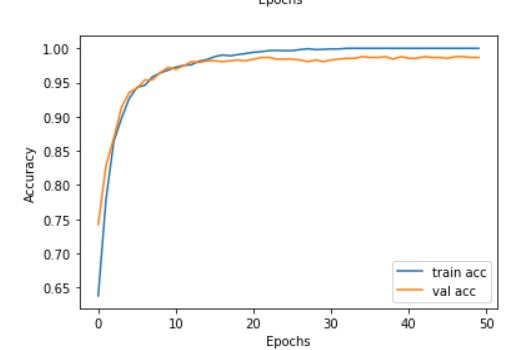
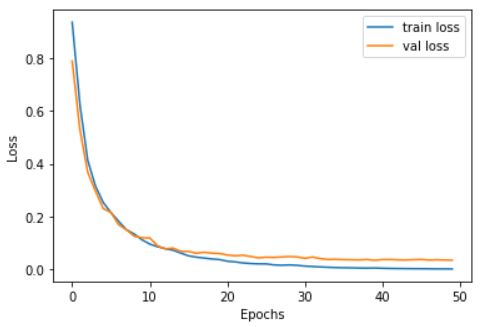


Fig 4 a) Accuracy and Loss Curves for 50 epochs

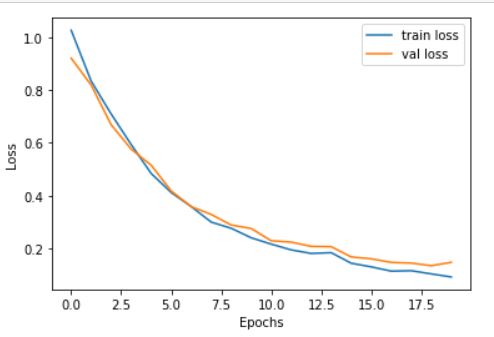
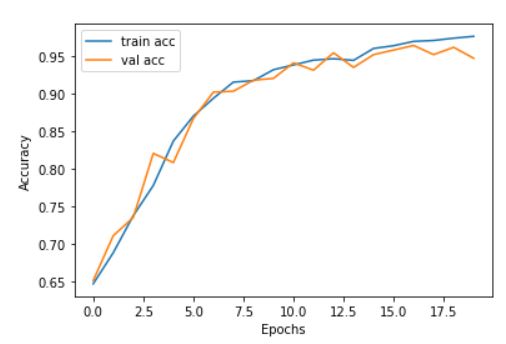
**** ****

Fig 4 b) Accuracy and Loss Curves for 20 epochs

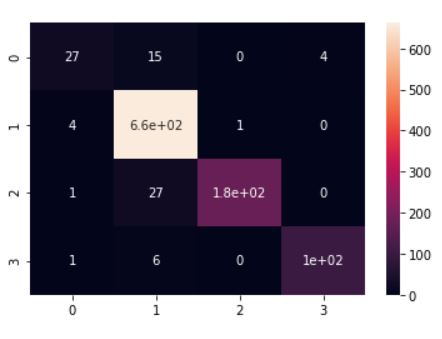


Fig 4 c) Confusion Matrix

**4.1 Analysis of Result**

For training of our model first, we took 50 epochs with batch\_size of 128 the training accuracy that we got was around 100 % with the loss of 0.0027 and validation accuracy of 98.66 % with validation loss of 0.0359 here we observed that after a particular number of epoch there was no increase in the accuracy score and there was no decrease in our loss function. Hence we did early stopping of our model.

Training our model with 50 epochs will make no sense if there is no change in our parameters hence we decided to decrease our epochs to 20. After that, we got training accuracy of 97.21% and validation accuracy of 94.95% with the loss of 0.0377 and 0.0978 for training and validation respectively.

And test the accuracy of 94.26% which seems that our model is performing well and is a generalized one.

Also by analyzing the confusion matrix we can say that we have few misclassified classes in our result.

**5.Conclusions**

In this project, we have implemented a Convolutional Neural network and validated our model with the accuracy metric. Also, we observed we can take few epochs to achieve a good result simply increasing the number of epochs does not mean that the model will give us a better result. Here we can now classify our images into different categories according to their species or body type from that we can study the characteristics of that particular bee which will help us to improve the health of the beehives and keep the bees out which damages or harm the beehives. Here we can also improve pollination and give importance to the bees which are contributing to improving the health of the hive with the help of this we can achieve a good quality of fruits, vegetables and other food items which can be very much beneficial for our community.

**6.References**

a) X. Wang, C. Chen, Y. Cheng, et al, Zero-shot image classification based on deep feature extraction.

United Kingdom: IEEE Transactions on Cognitive & Developmental Systems, 10(2), 1–1 (2018).

b) Shima Y. Image augmentation for object image classification based on the combination of pre-trained CNN and SVM. International Conference on Informatics, Electronics and Vision & 2017, an International Symposium in Computational Medical and Health Technology. 2018:1–6.

c) M.Z. Afzal, A. Kölsch, S. Ahmed, et al., Cutting the error by half: investigation of very deep CNN and advanced training strategies for document image classification

(Iapr international conference on document analysis and recognition. IEEE Computer Society, Kyoto, 2017), pp. 883–888.

d) Z.Yan, V.Jagadeesh, D.Decoste, et al., HD-CNN: hierarchical deep convolutional neural network for image classification.

Eprint Arxiv 4321-4329 (2014).

e) S. Roychowdhury, J. Ren, Non-deep CNN for multi-modal image classification and feature learning: an Azure-based model

(IEEE international conference on big data. IEEE, Washington, D.C., 2017), pp. 2893–2812.

f) Zhe Zhu, Dun Liang, Songhai Zhang, Xiaolei Huang,

Baoli Li, and Shimin Hu. Traffic-sign detection and classification in the wild.

In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

g) Pierre Sermanet, David Eigen, Xiang Zhang, Michael

Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization

, and detection using convolutional networks, 2013.