# DESIGN AND IMPLEMENTATION OF FUSIONET - A HYBRID MODEL TOWARDS IMAGE CLASSIFICATION

A Project Presentation by: Molokwu Reginald Chukwuka (2016 224 001)

Department of Computer Science, Faculty of Physical Science, Chukwuemeka Odumegwu Ojukwu University, Uli.

October 24, 2021

# Introduction

According to Goodfellow et al. (2017), Deep Learning is an approach to machine learning that has drawn heavily on our knowledge of the human brain, statistics and applied mathematics as it developed over the past several decades, and this approach aims at allowing computers to learn from experience and understand the world in terms of a hierarchy of concepts, with each concept defined in terms of its relation to simpler concepts. Thus by gathering knowledge from experience, this approach avoids the need for human operators to formally specify all of the knowledge that the computer needs. The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones. Therefore, if we decide to draw a graph showing how these concepts are built on top of each other; the graph is deep, with many layers. For this reason, the name Deep Learning (Deep Learning (DL)) was coined out.

# Statement of Problem

A major hassle of Artificial Intelligence (AI) is to devise effective and efficient means of transferring humans' informal knowledge (like sense of image recognition, sense of speech, etc.)into machines and computers such that these machines can act and behave exactly like humans. However, the occurrence of objects with respect to image representations in real world is usually associated with various features of variation or factors of influence which constitute distractions or noise in the image representations. Hence, it tends to be very difficult to actually disentangle these abstract factors of influence from the principal object or observed entity.Thus, an effective Convolutional Neural Network (CNN) model should be able to identify and focus on the principal object we aim to observe; and disregard the associated distractions (features of variation). To that effect, these remain open problems and challenges to CNN and modern AI.

# Objective of the Study

The goals of this study are:

- Proposition of a DL-based and hybrid model, FUSIONET, modelled for image prediction and classification problems in image analysis.
- Extensive bench-marking results which are centered on classic objective functions used for standard classifiers.
- Comparative analyses, between FUSIONET and state-of-the-art methodologies, against standard real-world image dataset.

# Scope of the Study

This study is limited to image recognition using CNN with respect to the research domain ofCV. Our baselines (datasets) for benchmarking the performance of our proposed model are asenumerated in Table 1

| Dataset | Type | Training set | Testing set | Classes |
|---|---|---|---|---|
| CINIC-10 | | 120,000 | 90,000 | 10 |
| CIFAR-10 | | 50,000 | 10,000 | 10 |
| CIFAR-100 | Multi-class | 50,000 | 10,000 | 100 |
| SVHN | | 73,000 | 60,000 | 10 |
| FoodNET-20 | | 1,500 | 900 | 20 |
| STL-10 | | 5,000 | 8,000 | 10 |

Table 1: Benchmark Dataset for Evaluation.

# Significance of the Study

The findings of this study will be most appealing to researchers and students in the domains of Computer Vision, Deep Learning, and Machine Learning. Optimistically, my work summarized herein will substantiate and advance recent applications of Deep Learning via Convolutional Neural Networks.

# Theoritical Review

Image classification is maybe the most significant piece of digital image analysis and computer vision Molokwu and Kobti (2019b). Image classification plays a pivotal role especially in face recognition, robot navigation, medical imaging and image search engines. With respect to advances in artificial intelligence (AI), real world (complex) images can be represented as vectors and analyzed by means of convolution neural network (CNN) operation. The exciting features of CNN is its ability to exploit spatial or temporal correlation in data LeCun et al. (1998).

# Theoritical Review Contd.

In recent time, various improvement geared towards CNN methodology have been proposed tomake CNN scalable to large heterogeneous, complex and multi class problems; this includes: modification of processing units Krizhevsky et al. (2012), He et al. (2016), Zeiler and Fergus(2014), Szegedy et al. (2016), parameter and hyper parameters Yoo (2019), Krishnakumari et al. (2020), Rawat and Wang (2019), Nguyen et al. (2019), optimization strategies Ismail et al. (2019), Ozcan and Basturk (2020), Dashdorj and Song (2019), Li et al. (2020) and design patterns Lu et al. (2020), Agarap (2017), Song et al. (2019), Suganuma et al. (2017).

# Theoritical Review Contd.

Molokwu (2019) A significant hassle in computer vision is to model an effective and efficient algorithm of transferring humans' informal knowledge into machines and computers such that these machines/computers can act and behave exactly like humans. However, the occurrence of objects concerning image representations in the real-world is usually associated with various features of variation or factors of influence that constitute distractions or noise in the image representations. Hence, it tends to be very difficult to disentangle these abstract factors of influence from the principal object or observed entity.

# Theoritical Review Contd.

To that effect, these remain open problems,and challenges to image classification, computer vision and machine learning. Herein our pro-posed methodology is based on an iterative learning approach which is targeted at solving the problems of image classification by combining 2 convolution operation models in parallel. primarily, learning in FUSIONET is induced via supervised training and FUSIONET is capable of learning the non-linear distributed features enmeshed in an image vector

# Analysis of the Present System

Image classification uses artificial intelligence technology to automatically identify objects,people, places and actions in images. One type of image classification algorithm is an image classifier. It takes an image (or part of an image) as an input and predicts what the image contains. The output is a class label, such as dog, cat or table. The algorithm needs to be trained to learn and distinguish between classes. This occurs by use of available ANN models and this models are very large in size because they have been trained on massive dataset, ex ImageNET. hence, scalability becomes a problem in that they require weeks of training on fast GPUs, sometimes months or years on training on Central Processing Unit (CPUs) for massive image datasets.

# Analysis of the Proposed System

Unlike a fully connected neural network, in a Convolutional Neural Network (CNN) the neurons in one layer don't connect to all the neurons in the next layer. Rather, a convolutional neural network uses a three-dimensional structure, where each set of neurons analyzes a specific region or "feature" of the image. CNNs filters connections by proximity (pixels are only analyzed in relation to pixels nearby), making the training process computationally achievable. In a CNN each group of neurons focuses on one part of the image. For example, in a cat image, one group of neurons might identify the head, another the body, another the tail, etc. There may be several stages of segmentation in which the neural network image recognition algorithm analyzes smaller parts of the images, for example, within the head, the cat's nose, whiskers, ears, etc. The final output is a vector of probabilities, which predicts, for each feature in the image, how likely it is to belong to a class or category.

# Analysis of the Proposed System Contd.

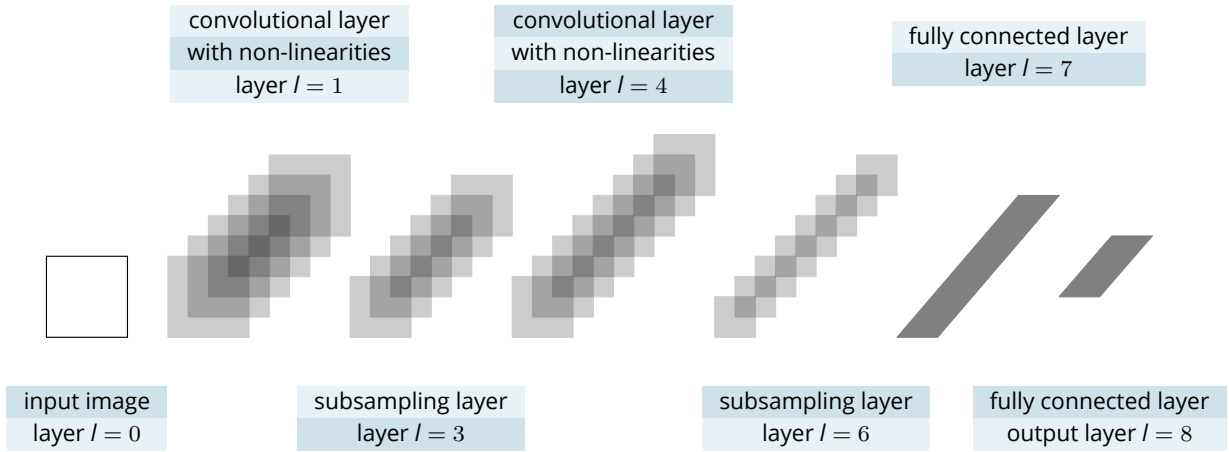


Figure 1: The architecture of the original convolutional neural network, as introduced by LeCun et al. (1989).

# Proposed Framework

Fig 2 illustrates the architecture of our proposition, FUSIONET

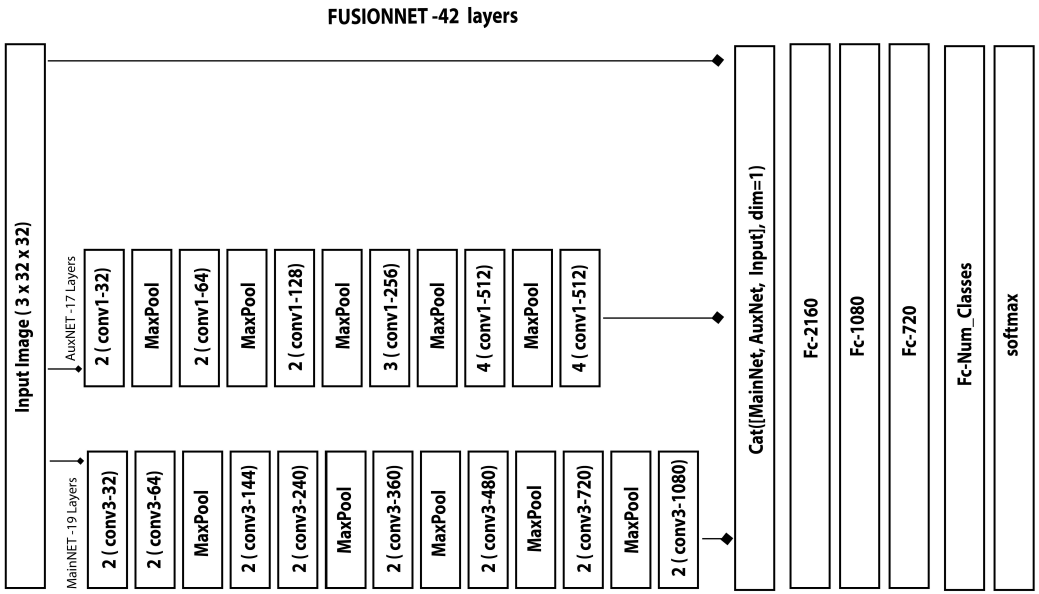


Figure 2: Proposed architectural framework of FUSIONET.

# Experiments, Results and Discussions

Table 2: Configuration of experimentation hyperparameters.

| | |
|---|---|
| Batch Size: 128 | Optimizer: SGD (Momentum=0.99) |
| Activation: PReLU | Epochs: 1000 |
| Dropout: 0.5 | Learning Rate: 0.003 |
| Learning Decay: after 300 epochs | weight decay: L2 penalty |
| Stride: 1 | L2 Multiplier: 5e-4 |
| Maxpooling: (2, 2) | Number of Parameters: 64M |

# Experiment

FUSIONET's experimentation setup was tuned in accordance with the hyperparameters shown in Table 2. Categorical Cross Entropy was employed as the cost/loss function; while the fitness was measured with reference to accuracy at *Top1, Top5 and Flops*. Moreover, the accuracy have been computed against each benchmark data set with regard to the constituent classes (or categories) present in each data set.

With regard to the image classification tasks herein, the performance of our FUSIONET model while benchmarking against five(5) popular baselines (NAT-M4, NAT-M3, SOPCNN, ResNeXt29x64d); and when evaluated against the validation/test samples for the benchmark data sets are as documented in Table 3, 4, 8, 6, 7

# Results

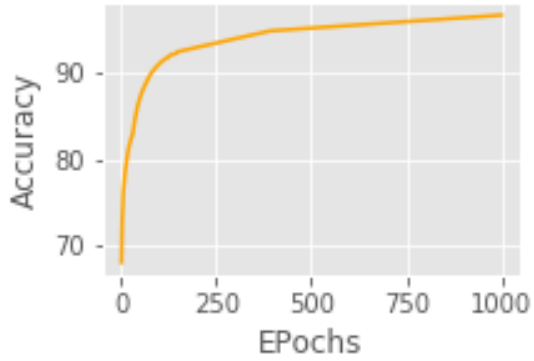


Figure 3: Epochs Vs Accuracy on CINIC-10 dataset after 1000 Epochs

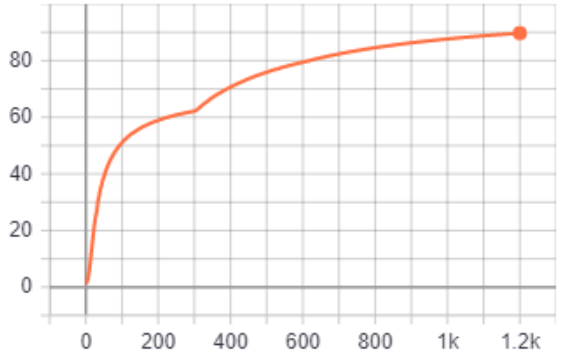


Figure 4: Epochs Vs Accuracy on CIFAR-100 dataset after 1200 Epochs

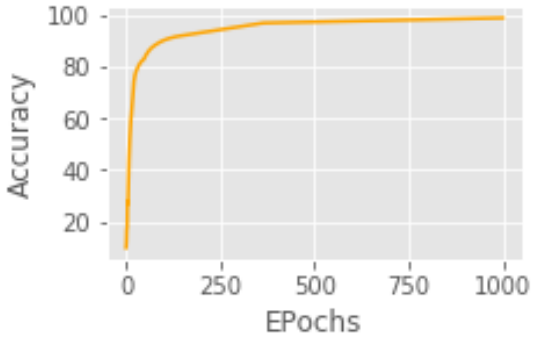


Figure 5: Epochs Vs Accuracy on CIFAR-10 dataset after 1000 Epochs

# Result Table

Table 3: Image classification over CINIC-10 data set. Results are based on the set apart validation samples

| CINIC-10 | | | | |
|---|---|---|---|---|
| Model | FLOPS | Top 1 accuracy | Top 5 accuracy | Error rate |
| NAT-M4 | 710M | 94.80% | 99.30% | 5.20% |
| NAT-M3 | **501M** | 94.30% | 98.60% | 5.70% |
| ResNeXt29x64d | 600M | 91.45% | 98.40% | 8.55% |
| VGG-16 | 690M | 87.77% | 97.55% | 12.23% |
| DenseNET-121 | **500M** | 91.26% | 98.53% | 8.74% |
| ResNET-18 | 505M | 90.27% | 98.01% | 9.73% |
| **FUSIONET** | 569M | **96.84%** | **99.88%** | **3.16%** |

# Result Table

Table 3. Shows our FUSIONET achieving state-of-the-art on the CINIC-10 dataset for top-1 and top-5 accuracy. Improving NAT-M4 by 2.04% for top-1 and 0.58% for top-5. floating point operations per second (FLOPS) was benched at 569M with regards to NAT-M4 of 710M.

# Result Table

Table 4: Image classification over CIFAR-10 data set. Results are based on the set apart validation samples

| CIFAR-10 | | | | |
|---|---|---|---|---|
| Model | FLOPS | Top 1 accuracy | Top 5 accuracy | Error rate |
| NAT-M4 | 468M | 98.40% | 99.60% | 1.60% |
| NAT-M3 | 392M | 97.20% | 98.30% | 2.80% |
| ResNeXt29x64d | 488M | 96.71% | 98.40% | 3.29% |
| SOPCNN | 440M | 94.29% | 98.00% | 5.71% |
| **FUSIONET** | **386M** | **98.54%** | **99.82%** | **1.46%** |

# Result Table

Table 5: Image classification over STL-10 data set. Results are based on the set apart validation samples

| STL-10 | | | | |
|---|---|---|---|---|
| Model | FLOPS | Top 1 accuracy | Top 5 accuracy | Error rate |
| **NAT-M4** | 573M | 92.61% | **98.55%** | 7.39 |
| NAT-M3 | **436M** | 97.80% | 98.48% | 2.20% |
| ResNeXt29x64d | 476M | 80.61% | 96.40% | 19.39% |
| SOPCNN | 424M | 88.08% | 97.02% | 11.92% |
| FUSIONET | 440M | **97.90%** | 98.76% | **2.10%**% |

# Result Table

Table 6: Image classification over SVHN data set. Results are based on the set apart validation samples

| SVHN | | | | |
|---|---|---|---|---|
| Model | FLOPS | Top 1 accuracy | Top 5 accuracy | Error rate |
| **NAT-M4** | 610M | **98.51%** | **99.67%** | **1.49%** |
| NAT-M3 | 520 | 97.80% | 99.24% | 2.20% |
| ResNeXt29x64d | 533M | 98.50% | 99.48% | 1.50% |
| SOPCNN | 512 | 98.50% | 99.55% | 1.50% |
| FUSIONET | **494M** | 97.63% | 99.31% | 2.37% |

# Result Table

Table 7: Image classification over CIFAR-100 data set. Results are based on the set apart validation samples

| CIFAR-100 | | | | |
|---|---|---|---|---|
| Model | FLOPS | Top 1 accuracy | Top 5 accuracy | Error rate |
| **NAT-M4** | 796M | 88.30% | **95.71%** | 11.70% |
| NAT-M3 | 492M | 87.70% | 95.55% | 12.30% |
| ResNeXt29x64d | 491 | 83.44% | 94.48% | 16.56% |
| SOPCNN | 501 | 72.96% | 90.22% | 27.04% |
| **FUSIONET** | **470M** | **89.72%** | **96.86%** | **10.28%** |

# Result Table

Table 8: Image classification over FoodNET-20 data set. Results are based on the set apart validation samples

| FoodNET-20 | | | | |
|---|---|---|---|---|
| Model | FLOPS | Top 1 accuracy | Top 5 accuracy | Error rate |
| NAT-M4 | 443M | 82.61% | 94.55% | 17.39 |
| NAT-M3 | **366M** | 87.80% | 94.48% | 12.20% |
| ResNeXt29x64d | 416M | 70.61% | 88.40% | 29.39% |
| SOPCNN | 394M | 78.08% | 87.02% | 21.92% |
| **FUSIONET** | 370M | **91.90%** | **98.76%** | **8.10%**% |

# Discussion

In this manner, we have highlighted the model which performed best (considering *Top-1, Top-5* precision and FLOPS where conceivable), for each classification task using a bold font. We utilized a point-based reviewing standard to discover the fittest model for each image classification task. The model with the most noteworthy combined point altogether wins the fittest model for each image classification task. Accordingly, as can be seen from our tabular results, FUSIONET is at the top with the highest fitness points; and this superb performance can be linked to two (2) primary factors,namely:

- The combination of 2 distinct convolution stack (MainNET and AuxNET) in FUSIONET conception model
- The top notch data pre-processing techniques employed herein with respect to the benchmark datasets. We ensured that all images used for training were in the standard format and classes.

# Summary

Machine learning with deep layer artificial neural networks (DL)has been gaining momentum over last decades. The successful results gradually propagate into our daily live. Convolutional neural networks (CNN) is a special architecture of artificial neural networks, proposed by LeCun et al. (1989). CNN uses some features of the visual cortex. One of the most popular uses of this architecture is image classification. Image classification refers to a process in computer vision that can classify an image according to its visual content. For example, an image classification algorithm may be designed to tell if an image contains a human figure or not.While detecting an object is trivial for humans, robust image classification is still a challenge in computer vision applications

# Conclusion

We propose a deep layer mathematical learning algorithm for real time image classification,which employs the use of convolutional neural network CNN. CNN is more contrast robust and overcomes the limitation of locality of multi-layer perceptron MLP. We combined the CNN with MLP classifiers in a novel approach to classify images in the dataset as seen in table 2

# Recommendation

- With the increase trend in image classification, Tertiary Schools in Nigeria should begin to adopt the scope of Image recognition for attendance management and exam impersonation discovery.
- The Federal Government of Nigeria, should invest in further research of the field of Machine learning, being a new field with little experts and broad scope.
- Industries in Nigeria, should begin to adopt image classification algorithms for object labeling, recognition and detection
- Deep layer Networks with automatic feature extraction have proven to be more efficient and time saving in the field of image classification, therefore should be widely used in image classification tasks rather than manual feature extraction.

# Further Research

In conclusion, we aspire to expand FUSIONET scope to include more computer vision problems. Also we are sourcing for additional benchmarking models and real world (complex) datasets for exhaustive validation of FUSIONET.