iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** V **Month of publication:** May 2019

DOI: https://doi.org/10.22214/ijraset.2019.5556

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

Object Detection and Visual Innovation using AR

Vaishnavi Bhat Marakini[1], Rahul Bhat[2], Meghavarshini M[3]
[1, 2, 3]Department of Information And Science, Global Academy of Technology, Bengaluru-560098

Abstract: This paper presents the implementation of applications that can detect objects and display information regarding it using reverse image search and various other AR applications. It discusses mainly about MobileNets for mobiles to detect an object. This MobileNets are based on streamlined architecture that uses depth-wise separable convolutions to build light weighted deep neural networks. This paper demonstrate the effectiveness of MobileNets across a wide range of applications and use cases of object detection.. This paper also talks about augmented reality, it is a technology which combines virtual objects and real-world environment. This paper proposes application for trying different furniture items in virtual way, measuring a plot and visualising it on 3D to place furnitures. It will eliminate the need of physically visiting the furniture store which is very time consuming activity.

***Index Terms*: Augmented reality, Convolutions, MobileNets.**

I. INTRODUCTION

Convolution neural network (CNN) could be a category of deep neural networks, most typically applied to analyzing visual imaging. MobileNets area unit supported efficient design that uses depth-wise severable convolutions to make lightweight weighted deep neural networks. CNNs also are called shift invariant or area invariant artificial neural networks (SIANN), supported their shared-weights design and translation unchangeability characteristics CNNs use comparatively bit of pre-processing compared to different image classification algorithms.

The general trend has been to form deeper and a lot of difficult networks so as to realize higher accuracy . However, these advances to boost accuracy aren't essentially creating networks a lot of economical with relation to size and speed. In several world applications like artificial intelligence, self-driving automobile and increased reality, the popularity tasks got to be dole out in an exceedingly timely fashion on a computationally restricted platform. increased Reality (AR) is employed to leverage the multiplied computing power of sensible phones and different visual interface to make a system wherever the user United Nations agency scans the objects through the phone or a visible interface enters the increased reality section and might simulate or move the entities to represent a virtual world and also the moves area unit as per the user's selection. In period if we have a tendency to users needed to shop for a piece of furniture objects while not visiting the outlets it had been attainable however it had been uphill to envision however the article truly appearance in home structure. The planned system with increased reality that lets user to do on virtual furniture to user's real home structure before shopping for from that user are going to be able to opt for furniture a lot of easier. the aim of the project is to develop a windows application for attempting totally different piece of furniture in virtual manner. it'll eliminate the requirement of physically visiting the piece of furniture store that is incredibly time intense activity

II. LITERATURE REVIEW

There has been rising interest in building little and economical networks within the recent literatures [5].They have designed a compendious and economical framework that may propose fewer candidate regions and extract a lot of discriminative options. That framework consists of 2 eight-layer CNNs that ar neat and powerful. To use CNNs to discover inshore ships, image samples ar needed, every of that ought to contain just one ship.

Real-Time Detection and following for increased Reality on Mobile Phones [1] .In which they need projected associate approach supported heavily changed progressive feature descriptors, specifically SIFT and Ferns and a template-matching-based huntsman exploitation 6DOF and SIFT and Ferns for real time detection and following for increased reality .

Shape Recognition and create Estimation for Mobile increased Reality [2].In which the system permits shapes that carry discourse meanings for humans to be used as increased Reality (AR) following targets. In time the user will teach the system for brand new poses .The system will acknowledge different shapes by learning antecedently trained inputs .

Touch-less interactive increased reality game on vision-based wearable device [3]. within which they need mentioned regarding

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

scrutiny analysis, that is projected to demonstrate the social acceptableness and usefulness of the touch-less approach, running on a hybrid wearable framework or with Google Glass, also as employment assessment. of these tools and visual innovations ar the new trends in AR. Marker primarily based increased Reality[4] This paper provides a comprehensive study of AR as well as its history, design, applications, current challenges and future trends.

International Journal of Advance Engineering and analysis Development[6]. during this paper they need provided summary of the fundamental theories and varied techniques connected with image mining. i.e. there's not one technique that provides in best for all user's necessities, here inventions of latest methodologies per necessities are going to be increase.

Recent Advance in Content-Based image retrieval[7]. during this paper, they need investigated the advance on content primarily based image retrieval in recent years. they need target the 5 key modules of the final framework, i.e., question formation, image illustration, image compartmentalization, retrieval marking, and search re-ranking. for every element, they need mentioned the key issues and categorized a spread of representative methods and strategies. Further, we've summarized eight potential directions which will boost the advance of content primarily based image retrieval within the close to future

III. METHODOLOGY

A. *Object Detection Using Convolution Neural Network*

The core layer of MobileNet is depth wise dissociable filters, named as depth wise dissociable Convolution. The network structure is another issue to spice up the performance. Finally, the dimension and determination are often tuned to trade off between latency and accuracy.

Depth wise dissociable convolutions that may be a sort of solved convolutions that factorize a customary convolution into a depth wise convolution and a 1×1 convolution that is known as as a degree wise convolution. In MobileNet, the depth wise convolution applies one filter to every input channel. the purpose wise convolution is then applied to a 1×1 convolution to mix the outputs of the depth wise convolution.

Assuming stride one artifact, the output feature of normal convolution is computed as:

$$G_{k,l,n} = \sum\_{i,j,m} K_{i,j,m} \cdot F_{k+i-1,1+j,m} \qquad (1)$$

One filter per input channel of Depthwise convolution can be written as:

$$\hat{G}_{k,l,m} = \sum_{i,j} \hat{K}_{i,j,m} \cdot F_{K+i-1,1+j-1,m} \qquad (2)$$

where K^ is that the depthwise convolutional kernel of size DK × DK × M wherever the mth filter in K^ is applied to the mth channel in F to supply the mth channel of the filtered output feature map G^ .

Depthwise convolution is expeditiously relative to plain convolution. but it filters solely input channels, it doesn't mix input channels to form new options. thus a further layer that computes a linear combination of the output of depthwise convolution via one × one convolution is required so as to come up with these new options.

The combination of depthwise convolution and one × one (pointwise) convolution is termed depthwise dissociable convolution that was originally introduced in .

which is that the total of the depthwise and one × one pointwise convolutions.

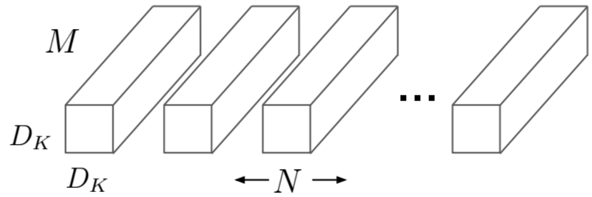
By expressing convolution as a 2 step method of filtering and mixing we have a tendency to get a discount in computation of:

$$\frac{D_k \cdot D_k \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F}{D_k \cdot D_k \cdot M \cdot N \cdot D_F \cdot D_F} = \frac{1}{N} + \frac{1}{D_k^2}$$

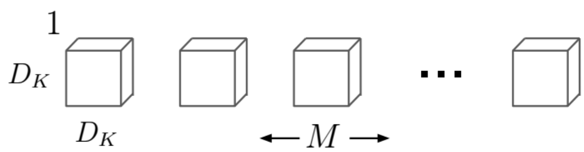
MobileNet uses three × three depthwise dissociable convolutions that uses between eight to nine times less computation than customary convolutions at solely alittle reduction in accuracy.

Additional factorization in spatial dimension like in doesn't save abundant further computation as little or no computation is spent in depthwise convolutions.

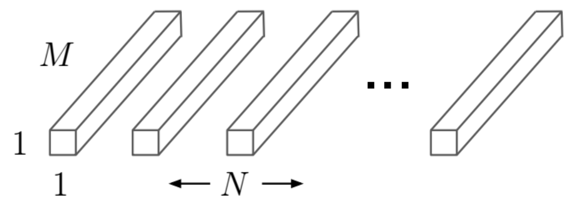
The following figure illustrates the distinction between customary convolution and depth wise dissociable convolution.



(a) Standard Convolution Filters



(b) Depthwise Convolutional Filters



(c) 1×1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

Depth wise separable convolution costs

$$D_k \cdot D_K \cdot M \cdot D_F \cdot D_F + M \cdot N \cdot D_F \cdot D_F$$

The following table shows the structure of MobileNet

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5× Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

The Width Multiplier is used to reduce the number of the channels. The Resolution Multiplier is used to reduce the input image of the network.

Table 4. Depthwise Separable vs Full Convolution MobileNet

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
Conv MobileNet	71.7%	4866	29.3
MobileNet	70.6%	569	4.2

Table 5. Narrow vs Shallow MobileNet

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
0.75 MobileNet	68.4%	325	2.6
Shallow MobileNet	65.3%	307	2.9

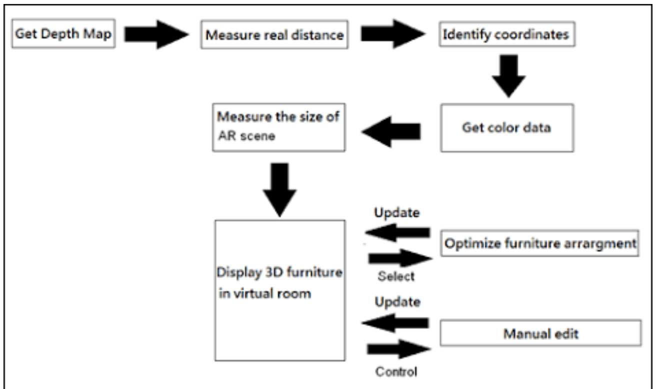
Table 6. MobileNet Width Multiplier

Width Multiplier	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
0.75 MobileNet-224	68.4%	325	2.6
0.5 MobileNet-224	63.7%	149	1.3
0.25 MobileNet-224	50.6%	41	0.5

Table 7. MobileNet Resolution

Resolution	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
1.0 MobileNet-192	69.1%	418	4.2
1.0 MobileNet-160	67.2%	290	4.2
1.0 MobileNet-128	64.4%	186	4.2

B. Placing Furniture Using Augmented Reality



The flow of the proposed system

The above figure shows how the system flows. When the system start, the system is going to get the depth and the color data , and measure a real distance and identify the coordinate for estimation of the wall skirting.

The floor size is estimated from the depth information and the furniture is scaled to real dimensions according to distance with respect to the camera.

User will choose the furniture, what they need from the furniture list and place them into the AR scene. Then, the user will depend upon their plan to manual edit the position of piece of furniture. In further, the system will mechanically prepare the chosen piece of furniture by prioritizing their shapes. Finally, user will see however the fashion is when place the piece of furniture in there and save the result for future reference.

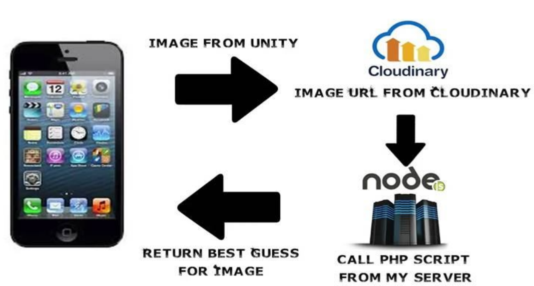
International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

Placing 3d Furniture At 3dplace

View AR software system was accustomed choose templates for the piece of furniture app. the bottom linear unit fixation and localisation was done and axes were mark. The 3D area is detected and therefore the catalogue from wherever the furniture's is born were additional to the info. For the empty plot feature the app initial localised the bottom or base and therefore the plot measure is finished victimization the ruler victimization AR feature. The app then takes the peak of the virtual building wall, the position of the door and therefore the window and displays the 3D read. The catalogue of furniture info seems which may be use to see if the furniture fits within the area well. View ar sdk provides varied tools and options to develop the appliance and therefore the info.

C. Reverse Image Search.

Object when scanned displays the information about the detected object .This is done through reverse image search. Unity's WWW class is the first step in doing Reverse search because it allows user to make requests via http to a server. When user visit a website search an information about an image, then user will type a url into his browser and making an http request to the server at that address. The web server returns the html code that your browser uses to render the webpage.
Then we need to make a request to www.unity.com.
After which we need to start a fresh unity project, create a C# script called test and drag it onto main camera.
First need to create an Enumerator called request, we want to work with www objects inside a co routine so we don't halt the entire program while things are processing. Then make sure that this code will run by starting the co routine in the start function.
Then string is created and defined as Web address to Unity inside the co routines.
Then a new www object is created and url is passed to it.
Yield returns www.Then string called html is created an defined it as www.text, which a string of text that was returned from the request.
Then actual API requests are made. By using Google custom search API we start searching. we need a unique key .Hence we search Google custom search API key. Then enable the custom search API for this app and copy the key to your clipboard.The image detected through Unity is sent to Cloudinary as an array of bytes. It processes it and returns the URL where the picture is hosted. From this we extract the title and run the php code in our server to display the best guess and the output. Problem that was encountered was, Google reverse image search gets redirected multiple times and finally is rendered as html via javascript. So, getting the data needed is not possible in Unity alone. The only way is to create api of sorts in php. Then pass the url of the image to search from Unity to a php script on the hosted server. The php script follows the redirects from Google using cURL and is able to view the javascript rendered html using Node.js



International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue V, May 2019- Available at www.ijraset.com

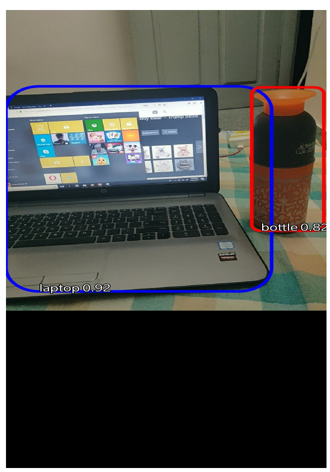
IV. FUTURE ENHANCEMENT

An area that may be revolutionized by AR is that the retail sector, wherever the technology can breach the gap between on-line and physical outlets, providing a richer expertise and personalized promotions delivered right to the user. laptop games area unit clearly another vital space of development for AR.

The mobile business is additionally a possible scene for the longer term of AR. From good phones accessing GPS to wearable technologies starting from Google Glass to Apple Watch, increased reality are going to be expressed by specific apps and micro-location technologies like iBeacons. All can eventually blur the excellence between reality and virtuality – delivery users a number of relevant information to fully alter their expertise. There area unit already apps which provide some basic versions of this practicality on Smartphone. another potential future uses of increased reality area unit within the sector of drugs (direct coaching of students), military (precise location of enemy positions), creation (mixed media and interactive artworks), business (assistance in prototyping) and teaching (visualizing learning material in interactive apps).

V. RESULTS

Furniture AR



Object detection using tensor flow

Ruler using AR

Information displayed through reverse image search

VI. CONCLUSION

This paper projected a brand new model design known as MobileNets supported depthwise severable convolutions. we have a tendency to investigated a number of the vital style choices resulting in AN economical model. we have a tendency to then incontestable however MobileNets use breadth number and backbone number by commercialism off an inexpensive quantity of accuracy to scale back size and latency. we have a tendency to complete by demonstrating MobileNet's effectiveness once applied to a good sort of tasks. As a next step to assist adoption and exploration of MobileNets, we have a tendency to arrange on emotional models in Tensor Flow.

This paper projected a system which can facilitate the client to look at the furnishings object just about in real setting before shopping for the item. because of these system client can come back to understand however his home structure would take care of shopping for the furnishings object. These projected system would let the user to do multiple combination of object just about while not physical movement of furnishings objects. These can facilitate the client to work out a way to setup furnishings in home structure

REFERENCES

[1] Daniel Wagner," Real-Time Detection and Tracking for Augmented Reality on Mobile Phones ", IEEE transactions on visualization and computer graphics, vol. 16, no. 3, May/June 2010.

[2] Nate Hagbi, "Shape Recognition and Pose Estimation for Mobile Augmented Reality", IEEE transactions on visualization and computer graphics, vol. 17, no. 10, october 2011.

[3] Zhihan Lv, Alaa Halawani, Shengzhong Feng, Shafig Ur , Haibo Li , "Touch-less interactive augmented reality game on vision-based wearable device", Journal Personal and Ubiquitous Computing Vol. 19, no. 3, July 2015.

[4] Anuroop Katiyar, Karan Kalra2and Chetan Garg," Marker Based Augmented Reality", Advances in Computer Science and Information Technology, vol. 2, no. 5, April-June, 2015.

[5] Xiaobin Li, Shengjin Wang,"Object Detation using Convolutional Neural Networks in a Coarse-to-Fine Manner", IEEE Geoscience and Remote Sensing Letters , vol 14 ,Issue 11, Nov 2017 .

[6] Lad Ami D. , Mori Shilpa I. , Raulji Urvashi K.. , Shaikh Saliha H." International Journal of Advance Engineering and Research Development", vol 4, Issue 4, April 2017.

[7] Wengang Zhou, Houqiang Li, and Qi Tian "RecentAdvance in Content-Based image retrieval: A literature survey",arxiv, sept 2017.