

**THE REPUBLIC OF TURKEY
BAHCESEHIR UNIVERSITY**

**IMPACT OF QUALITY DISTORTIONS ON CNN IMAGE
CLASSIFIERS AND RESTORATION USING DEEP
CONVOLUTIONAL AUTO-ENCODER**

Master's Thesis

ÇAĞKAN ÇİLOĞLU

ISTANBUL, 2018

**THE REPUBLIC OF TURKEY
BAHCESEHIR UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED
SCIENCES
COMPUTER ENGINEERING**

**IMPACT OF QUALITY DISTORTIONS ON CNN
IMAGE CLASSIFIERS AND RESTORATION
USING DEEP CONVOLUTIONAL AUTO-
ENCODER**

Master's Thesis

ÇAĞKAN ÇİLOĞLU

Supervisor: PROF. DR. TAŞKIN KOÇAK

ISTANBUL, 2018

**THE REPUBLIC OF TURKEY
BAHCESEHIR UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
COMPUTER ENGINEERING**

Name of the thesis: Impact of Quality Distortions on CNN Image Classifiers and Restoration Using Deep Convolutional Auto-Encoder

Name/Last Name of the Student: Çağkan Çiloğlu

Date of the Defense of Thesis: 25/05/2018

The thesis has been approved by the Graduate School of Natural and Applied Sciences.

Assist. Prof. Yücel Batu SALMAN
Graduate School Director
Signature

I certify that this thesis meets all the requirements as a thesis for the degree of Master of Science.

Assist. Prof. Tarkan AYDIN
Program Coordinator
Signature

This is to certify that we have read this thesis and we find it fully adequate in scope, quality and content, as a thesis for the degree of Master of Science.

Examining Committee Members

Signature

Thesis Supervisor
Prof. Dr. Taşkın KOÇAK

Member
Assoc. Prof. Dr. Songül Varlı ALBAYRAK

Member
Assist. Prof. Tarkan AYDIN

ABSTRACT

IMPACT OF QUALITY DISTORTIONS ON CNN IMAGE CLASSIFIERS AND RESTORATION USING DEEP CONVOLUTIONAL AUTO-ENCODER

Çağkan ÇİLOĞLU

Computer Engineering Graduate

Thesis Supervisor: Prof. Dr. Taşkın KOÇAK

May 2018, 42 pages

Deep neural networks have obtained significant performance on recognizing objects even in real-time video stream. As hardware requirements of this task decrease in cost and parts getting smaller, this technology can be used in mobile devices. Most of the time, deep neural networks are trained and tested on quality picture datasets. But frames provided by camera may be distorted because of camera defects and/or weather conditions such as rain and snow. This paper provides an evaluation of four state-of-the-art deep neural network models for picture classification under quality distortions. Three types of quality distortions are considered: blur, noise and contrast. It is shown that the existing networks are susceptible to these quality distortions and architecture of the network dramatically affects the results. Using deep convolutional auto-encoder to restore picture quality is suggested and better scores have been archived utilizing it. Results enable future work in developing machine vision systems on that are more invariant to quality distortions.

Keywords: Convolutional Neural Network, Image Classification, Image Restoration Auto-Encoder, Object Classification in Video, iOS CoreML

ÖZET

KALİTE BOZULMALARININ SINIFLANDIRICI CNN'LER ÜZERİNDEKİ ETKİLERİ VE DERİN KONVOLÜSYONEL AUTO-ENCODER İLE DÜZELTİLMESİ

Çağkan ÇİLOĞLU

Bilgisayar Mühendisliği Yüksek Lisans

Tez Danışmanı: Prof. Dr. Taşkın KOÇAK

Mayıs 2018, 42 sayfa

Derin sinir ağları gerçek zamanlı video akışında bile nesnelere tanıma üzerinde önemli başarı elde etmiştir. Bu görevin donanımsal gereksinimlerinin maliyetinin azalması ve fiziksel olarak küçülmesi bu teknolojinin mobil cihazlarda kullanılabilmesini sağlamıştır. Çoğulukla, derin sinir ağları kaliteli veri setleri üzerinde eğitilmiş ve test edilmiştir. Fakat kamera tarafından sağlanan kareler, kamera kusurları ve/veya hava koşulları nedeniyle bozulmuş olabilir. Bu tez, resmin kalitesinin bozulma durumu altında, resim sınıflandırması için kullanılan, yüksek performanslı en güncel dört derin sinir ağ mimarilerinin bir değerlendirmesini sunar. Üç tür kalite bozulması üzerinde çalışılmıştır: bulanıklık, gürültü ve kontrast. Mevcut ağların bu kalite bozulmalarına duyarlı olduğu ve mimarilerinin sonuçlar üzerinde direkt etkisi olduğu gösterilmiştir. Bu durum karşısında resim kalitesini arttırmak için derin konvolüsyonel auto-encoder kullanılması önerilmiş ve bozuk resimlere göre daha doğru değerler alınmıştır. Bu tez, makine görü sistemlerinin resim kalitesinden bağımsız olarak sonuç vermesinin geliştirilmesine katkı sağlar.

Anahtar Kelimeler: Konvolüsyonel Sinir Ağları, Resim Sınıflandırma, Auto-Encoder ile Resim Düzeltme, Videoda Obje Sınıflandırılması, iOS CoreML

CONTENTS

FIGURES.	vii
ABBREVIATIONS	viii
SYMBOLS	ix
1. INTRODUCTION	1
1.1 BACKGROUND INFORMATION	2
2. LITERATURE REVIEW	5
3. DATA AND METHOD	7
3.1 DATASET	7
3.2 QUALITY DISTORTIONS	7
3.3 DEEP NEURAL NETWORK ARCHITECTURES	8
3.4 RESTORATION CONVOLUTIONAL AUTO-ENCODER	10
3.5 METHOD	12
3.5.1 Finding the Impact of Distortion Effects	12
3.5.2 Restoration Stage	13
4. FINDINGS	15
4.1 QUALITY DISTORTION EFFECTS	15
4.1.1 Blur	15
4.1.2 Noise	16
4.1.3 Contrast	17
4.2 RESTORATION	18
4.2.1 Blur	19
4.2.2 Noise	24
4.2.3 Contrast	30
5. DISCUSSION	35
5.1 QUALITY DISTORTIONS	35
5.2 RESTORATION CONVOLUTIONAL AUTO-ENCODER	37
6. CONCLUSIONS AND FUTURE WORK	40
REFERENCES	43

APPENDICES

Appendix A.1 Wooden spoon - Quality distortions impact tables.....	47
Appendix A.2 Washbasin - Quality distortions impact tables.....	48
Appendix A.3 Sunglasses - Quality distortions impact tables.....	49
Appendix A.4 Electric switch - Quality distortions impact tables.....	50
Appendix A.5 Envelope - Quality distortions impact tables.....	51
Appendix A.6 Toilet seat - Quality distortions impact tables.....	52
Appendix A.7 Plastic bag - Quality distortions impact tables.....	53
Appendix A.8 Soup bowl - Quality distortions impact tables.....	54
Appendix A.9 Band-aid - Quality distortions impact tables.....	55
Appendix A.10 Dishwasher - Quality distortions impact tables.....	56
Appendix A.11 Watch - Quality distortions impact tables.....	57
Appendix A.12 Pill bottle - Quality distortions impact tables.....	58
Appendix A.13 Toilet paper - Quality distortions impact tables.....	59
Appendix A.14 Remote controller - Quality distortions impact tables.....	60
Appendix A.15 Soap dispenser - Quality distortions impact tables.....	61
Appendix A.16 Banana - Quality distortions impact tables.....	62
Appendix A.17 Television - Quality distortions impact tables.....	63
Appendix A.18 Washing machine - Quality distortions impact tables.....	64
Appendix A.19 Coffee mug - Quality distortions impact tables.....	65
Appendix A.20 Pillow - Quality distortions impact tables.....	66

FIGURES

Figure 3.1: Restoration auto-encoder model architecture.....	10
Figure 3.2: Experiment setup.....	11
Figure 3.3: Raw and distorted pictures from average and quality cameras.....	13
Figure 4.1: Blur distortion effects.....	15
Figure 4.2: Noise distortion effects.....	16
Figure 4.3: Contrast distortion effects.....	17
Figure 4.4: Distorted and restored coffee mug.....	18
Figure 4.5: PSNR values in restoration of blur distortion.....	20
Figure 4.6: Blur restoration performance for Inception v3.....	20
Figure 4.7: Blur restoration performance for ResNet50.....	21
Figure 4.8: Blur restoration performance for SqueezeNet.....	22
Figure 4.9: Blur restoration performance for MobileNet.....	23
Figure 4.10: Distorted and restored remote controller.....	24
Figure 4.11: PSNR values in restoration of noise distortion.....	25
Figure 4.12: Noise restoration performance for Inception v3.....	25
Figure 4.13: Noise restoration performance for ResNet50.....	27
Figure 4.14: Noise restoration performance for SqueezeNet.....	28
Figure 4.15: Noise restoration performance for MobileNet.....	29
Figure 4.16: Distorted and restored wooden spoon.....	30
Figure 4.17: PSNR values in restoration of contrast distortion.....	31
Figure 4.18: Contrast restoration performance for Inception v3.....	31
Figure 4.19: Contrast restoration performance for ResNet50.....	32
Figure 4.20: Contrast restoration performance for SqueezeNet.....	33
Figure 4.21: Contrast restoration performance for MobileNet.....	34

ABBREVIATIONS

CAE	:	Convolutional Auto-Encoder
CNN	:	Convolutional Neural Network
DNN	:	Deep Neural Network
MB	:	Megabyte
PSNR	:	Peak Signal-to-Noise Ratio

SYMBOLS

Bias term	:	b
Decibel	:	dB
Input to the neuron	:	x
Maximum possible pixel value	:	MAX_I
Mean squared error	:	MSE
Nonlinearity function	:	φ
Weight vector	:	w

1. INTRODUCTION

Deep neural networks have obtained significant performance on recognizing objects even in real-time video stream. As hardware requirements of this task decrease in cost and parts getting smaller, this technology can be used in mobile devices. This development makes the quality of frames provided by camera an important concern.

Most of the time, deep neural networks are trained and tested on quality picture datasets. But frames provided by camera may be distorted because of camera defects and/or weather conditions such as rain and snow or user's movements. In this context it is important to know how quality distortions impact on results of image classification.

Because utilizing deep neural networks on mobile devices is a new development there are limited number of architectures that are optimized enough to run on mobile devices. The questions would be, are all these architectures effected by quality distortions equally, do distortion types effect the result all the same or some more than others, which levels of distortions significantly affect the classification results.

It is possible to restore the images damaged by these distortions using neural network. But as using neural networks on mobile devices is a new development it is vital to learn if using restoration convolutional auto-encoder(CAE) would be suitable in this environment. If it is, then the next question would be if the architecture of image classification model poses as a factor in the results or the image provided by restoration model gives the same output for all image classification architectures.

To answer these questions, an evaluation of four state-of-the-art deep neural network models for picture classification under quality distortions on mobile environment has been made. Three types of quality distortions have been considered: blur, noise and contrast. Impacts of these distortions on image classification results have been compared using different models. Suitability of using restoration CAE on mobile devices have been investigated. Then classification results of restored images have been assessed in terms

of image classification architectures, distortion types and levels. Lastly, results gotten by these experiments have been discussed and a conclusion has been given by stating a general understanding and suggestions for future works to develop machine vision systems on mobile devices that are more invariant to quality distortions.

In this work, CoreML framework on iOS has been used to handle the communication between the device and the neural network models. Except the training of the restoration auto-encoder, for all other procedures such as data gathering, assessing the performance of image classifiers and restoration auto-encoder model, only an iPhone 6 device has been used to limit the scope to mobile usage.

1.1 BACKGROUND INFORMATION

In computer vision, deep neural networks (DNN) achieves state-of-the-art results within numerous domains such as processing pictures, video, speech and audio (LeCun *et al.* 2015, pp.436-444).

DNNs are a set of algorithms, modeled after the human brain, which are designed to recognize patterns in given input. In common words, a deep network comprises of layers of neurons where each neuron computes function:

$$f(x) = \varphi (w^T x + b) \tag{1.1}$$

where x is the input to the neuron, w is a weight vector, b is a bias term and φ is a nonlinearity function. Each neuron gets possibly numerous inputs and yields a single number. The importance of nonlinearity is allowing layers of neurons to learn non-linear functions. In these layered designs, the output of one layer of units gets to be the inputs to the next layer of units.

For picture recognition tasks, the input to the network is the picture itself (most of the time normalized). However, in the event that a single neuron is to get inputs from the whole picture, the memory and computational necessities rapidly end up restrictive. Weight sharing is being used to solve this issue. Instead of each neuron using a different

weight vector w , it is shared between neurons. The weight vector behaving like a bridge connects to adjacent neurons from the previous layer inside a pre-defined region known as a receptive field. Layers with convolutional shared weights are defined as convolutional layers. Whereas layers without the convolutional shared weights are defined fully connected layers.

Convolutional layers together with pooling layers and fully connected layers compose convolutional neural network (CNN) (LeCun *et al.* 1998) a type of DNN. CNN architectures make the assumption that the inputs are pictures, which permits to use certain techniques into the architecture such as filtering to find local feature maps, pooling to reduce the dimensions of each feature map while preserving the significant features and weight sharing. These features make the architecture more efficient in terms of calculation by significantly reducing the number of parameters within the network. CNNs are used primarily for image processing (Ciresan *et al.* 2012; Szegedy *et al.* 2014), video analysis (Ji *et al.* 2012) and natural language processing (Collobert and Weston 2008).

A softmax layer is used as a last part of the network most of the time. Softmax normalizes the outputs by having their sum to one. By doing so the output layer functions as a probability distribution with each neuron corresponding to the probability of the network for a specific class.

The parameters w and b are trained with a huge set of input pictures. To begin with the output of the network is calculated for a given set of pictures. This value is compared to the known class tags then a cost function is used to verify the result by finding how close the prediction to ground truth is. The gradient of this cost function can be calculated, and by propagating this value in reverse through the network, each neurons gradient can be calculated also. With knowing the gradient, any optimization methods based on gradient descent can be used to optimize the weights. This technique is called the backpropagation algorithm (Rumelhard *et al.* 1986, pp. 533–536).

One specific domain of picture classification on extensive scale datasets containing many pictures from huge number of classes. These problem domains were thought to be greatly

difficult however CNNs have accomplished extremely great outcomes. Despite their great success, deep networks are susceptible to adversarial samples (Goodfellow *et al.* 2015). Adversarial samples are created by adding worst-case noise to the picture to such an extent that the classification prediction is inaccurate with a high confidence.

In processing real-time video stream from camera, the network may experience quality distortions. While using a mobile device to classify frames, user may shake his/her device or camera being out-of-focus would cause blur. Another possible distortion noise may result from acquiring frames with low quality camera sensors, in poor illumination conditions or in bad weather conditions. Although these distortions are not the worst case, they can in any case make the system misclassify pictures.

As well as picture classification, CNNs show potential in correcting quality distortions such as blur and noise (Fu *et al.* 2016; Gao and Grauman 2017). A type of CNN, convolutional auto-encoders(CAE) is state-of-the-art tool in restoring images. Auto-encoder is an unsupervised neural network based on an encoder-decoder template, in which an encoder transforms the input into a low-dimensional representation, and a decoder is trained to recreate the original input using this representation (Hinton and Salakhutdinov 2006). Main difference of CAEs from auto-encoders is that CAEs are using weight sharing and so preserving spatial locality (Masci *et al.* 2011). But it is yet to be studied if they are suitable to use on real-time video stream from a mobile device also the results they provide would have any effect on architectures optimized for mobile usage.

For these sorts of distortions, clarifying the impact of the distortion levels which influence the networks in such way that they begin to misclassify frames coming from device's camera is one of the two goals of this paper. Examining that if a restoration CAE can run on mobile device and if it can be used on real-time video stream to restore picture quality in terms of these distortions and comparing the results produced by undistorted pictures are together form the second goal. Results will enable future work in developing machine vision systems on mobile devices that are more invariant to quality distortions.

2. LITERATURE REVIEW

For most applications in computer vision it is expected that the input pictures are of generally without any distortion. But a camera attached on mobile device much likely will produce distorted frames as an input.

Many computer vision applications such as face-filters and face-swapping are already available for mobile phones as hardware requirements of the task decrease in cost. Next milestone will be the integration of deep neural network to these systems. There are researchers who proposed efficient CNNs for mobile devices (Zhang *et al.* 2017; Howard *et al.* 2017). This development makes the prerequisite of high quality pictures for deep neural networks a concern.

Before that distorted pictures as input was not considered as a problem because supplying quality images to CNNs was user's responsibility. Being a relatively a new concern behavior of CNNs under these conditions is still an active field of researchers.

Nguyen, Yosinski and Clune (2015, p.8) presented that discriminative DNN models are easily manipulated in such way that they labeled many unrecognizable images with near-certainty as members of a recognizable class. They produced images which are entirely unidentifiable to human observers, but that state-of-the art DNNs were certain with 99.99% confidence that they are recognizable objects.

Dodge and Karam (2017, p.5) showed that the performance of DNNs and human subjects is roughly the same on clean images, but for distorted images human subjects achieve much higher accuracy than DNNs. In other words, DNNs cannot effectively recognize and get the visual data with these kinds of distortions.

Zhou, Song and Cheung (2017, p.4) examined the CNN picture classifier performance under blur and noise distortions. Experimenting on CIFAR-10 (Krizhevsky 2009) and

ImageNet (Deng *et al.* 2009) datasets, they stated that these types of distortions significantly worsen the error rate because they affect the edge, color and texture information. They presented their fine-tuning and re-training the model using noisy data can increase the model performance on distorted data, and re-training method usually achieves comparable or better accuracy than fine-tuning.

Borkar and Karam (2018) proposed a metric to detect the most noise susceptible convolutional filters and sort them in order of the highest gain in classification accuracy when corrected. In their proposed approach named DeepCorrect, at the output of these sorted filters, they applied small stacks of convolutional layers with residual connections and trained them to correct the worst distortion affected filter activations, while leaving the rest of the pre-trained filter outputs in the neural network unchanged.

For image restoration, Mao, Shen and Yang (2016) suggested a very deep fully convolutional auto-encoder network, that is an encoding-decoding template with symmetric convolutional and deconvolutional layers. While convolutional layers grasping the abstraction of image with eliminating corruptions, deconvolutional layers recover the image details as they have the capability to up-sample the feature maps. As such deep neural network model as this would need long time to train they proposed to symmetrically link convolutional and deconvolutional layers with skip-layer connections, to have the training converge much faster and attain better results. They also showed it is possible to handle different levels of corruptions using a single model.

In this paper, first assessment of deep networks that are running on a mobile device on video frames under different types and diverse levels of picture quality distortions has been presented. In differentiate to other papers, real-time video stream has been used as input to deep networks that trained by high quality pictures. Frames has been augmented by introducing distortions on real-time and has been used to assess the performance of state-of-the-art image classifier deep neural networks. As a second task, restoring frames by utilizing a deep convolutional auto-encoder neural network model has been attempted and the results between restored frames, distorted ones and the original ones have been compared under different types of distortions and levels.

3. DATA AND METHOD

Only an iOS device has been used to gather the data, to distort frames and to utilize deep neural networks with this, scope has been limited to the device's capabilities.

3.1 DATASET

In this paper, only real-time frames have been used in the experiment. In total almost 360k frames from 20 different objects have been tested to understand the effects of image distortions. 3 of these objects have been used in restoration experiment. From household items to electronics, objects are selected from various categories.

Objects sizes and their features were also considered as an outcome objects such as band-aid and a big plastic bag are included to the dataset. Colors on the object was also considered, thus objects with single color such as a banana and multiple colors such as a TV remote have been added. Lastly, detailed objects such as wrist watch and more homogenous objects such as paper-towel have been selected.

3.2 QUALITY DISTORTIONS

Distortions were applied to every frame of the real-time stream directly with OpenGL filters. 3 types of image distortions have been used in the assessment of deep networks: blur, noise and contrast.

Blur is the result of camera not being focused properly on the object of interest. Furthermore, blur can mimic the network's execution on small or far off objects which will be shot with low resolution. Intensity of the blur from 1 to 10 has been used in tests.

Noise may result from utilizing low quality camera sensors. Strength of the noise from 10 to 100 in steps of 10 has been varied in the tests.

Contrast of the frames has been reduced as the final distortion test. Contrast decrease is gotten by mixing the input picture with a gray picture (Haeberli and Voorhies 1994, pp. 8-9). Level of contrast is indicated by the blending factor. The blending factor from 0 to 1 in steps of 0.1 has been varied in the tests.

3.3 DEEP NEURAL NETWORK ARCHITECTURES

In this paper four neural network architectures have been studied: Inception v3 (Szegedy *et al.* 2015), ResNet50 (He *et al.* 2015), SqueezeNet (Iandola *et al.* 2016) and MobileNet (Howard *et al.* 2017). They are lightweight enough to be utilized on mobile devices. All four of these networks have been trained on the ImageNet dataset. ImageNet is an image dataset which contains over 14 million images under approximately 20 thousand categories (ImageNet Official Website 2010). Images of each category are quality-controlled and human-annotated (Deng *et al.* 2009). Pre-trained model weights of these networks have been used in tests.

Inception model includes a sort of layer called inception layers. By processing the input with different size filters in parallel, the inception layers combine the filter responses together. In simple terms, the layer acts as multiple convolution filters. These filters are applied to the same input with some pooling then results are concatenated. This allows the model to take advantage of multi-level feature extraction as it extracts general (5x5) and local (1x1) features in parallel. Also, Inception structure enables the network to use far less parameters (Szegedy *et al.* 2014). Inception v3 is the optimized version for mobile vision and big-data scenarios.

By introducing “residual learning framework” (He *et al.* 2015, p.3) ResNet maintained its low complexity with a depth of up to 152 layers. In general usage, several layers are stacked and trained in a deep convolutional neural network. The network learns several low, mid and high level features at the end of its layers. In residual learning, network also tries to learn residuals so that predictions are close to the actuals. In numerical analysis a residual is the error in a result. For example, if the prediction of age of a person is 18 while her actual age is 20, result will be off by 2. In this result, 2 will be residual. ResNet

learns residuals by using shortcut connections. Shortcut connections skip one or more layers and they simply perform identity mapping. Then their outputs are added to the outputs of the stacked layers. A cluster of these residual nets accomplishes 3.57 percentage error on the ImageNet data set. ResNet50 is a 50-layer residual network. As it is lightweight enough ResNet50 a suitable deep network for mobile vision usage.

SqueezeNet is only 5 MB (megabyte) and utilizes 50 times less parameters than AlexNet but their accuracies are similar (Iandola *et al.* 2016, p.6). Architecture uses three main strategies for reducing parameters. First strategy is making the network smaller by replacing 3x3 filters with 1x1 filters to reduce the number of parameters 9x by replacing a bunch of 3x3 filters with 1x1 filters. Typically, a larger 3x3 convolution filter captures spatial information of pixels close to each other. On the other hand, 1x1 convolutional filters zero in on a single pixel and capture relationships amongst its channels as opposed to neighboring pixels. Next strategy is reducing the number of inputs for the remaining 3x3 filters. This strategy reduces the number of parameters by basically just using fewer filters. Last strategy is doing the downsample late in the network so that convolution layers have large activation maps. Having larger activation maps near the end of the network is in stark contrast to networks like VGG where activation maps get smaller as getting closer to the end of a network. Even though SqueezeNet is not the best network when it comes precise outputs, it's an extremely efficient deep network.

MobileNet is a light weight deep neural network architecture for mobile and embedded vision applications. Most distinctively, it uses depthwise separable convolutions. By using this type of convolutions, the number of parameters significantly reduces when compared to the network with normal convolutions with the same depth. In MobileNet normal convolution is replaced by depthwise convolution subsequently pointwise convolution which is called depthwise separable convolution. In another terms, architecture sacrifices accuracy for low complexity deep neural network. With that by using two global hyper-parameters that tradeoff between latency and accuracy, it is possible to choose the right sized model for the applications based on the constraints of the problem (Howard *et al.* 2017 p.1).

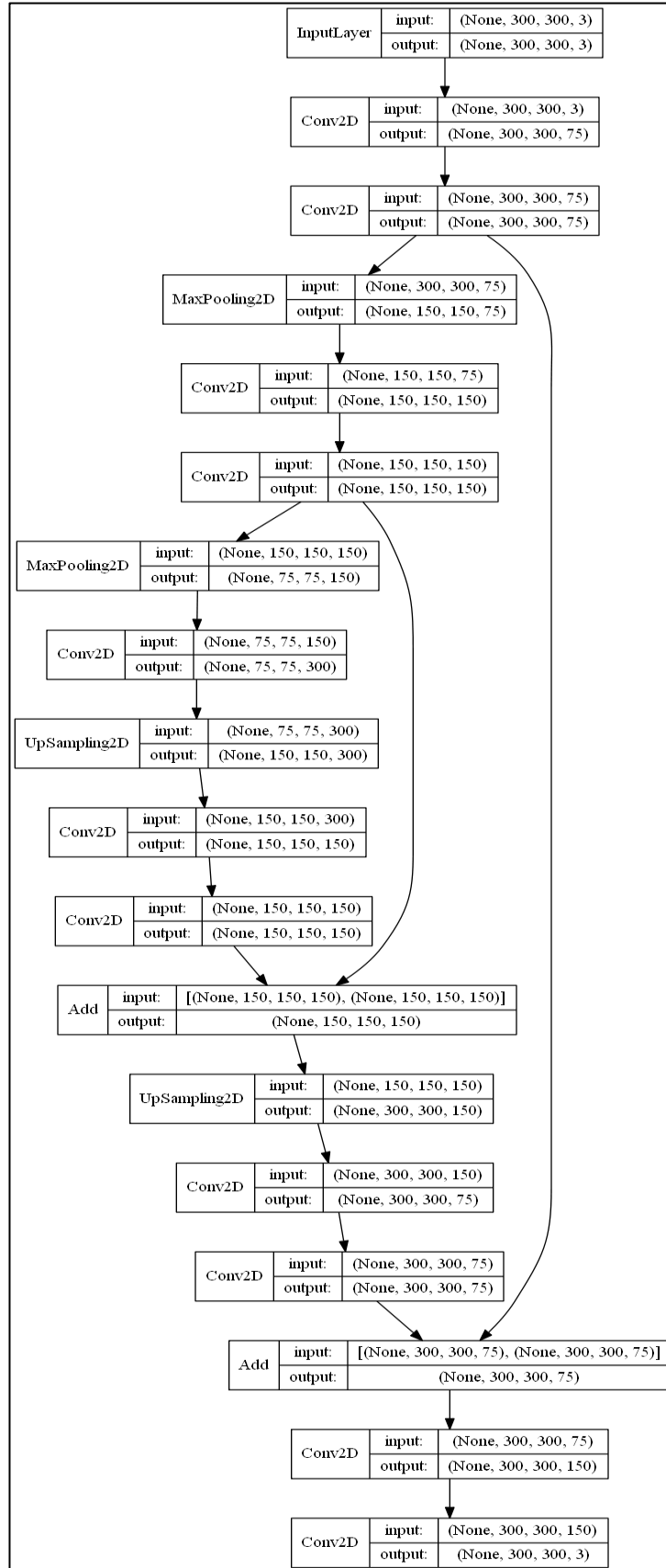
3.4 RESTORATION CONVOLUTIONAL AUTO-ENCODER

As second part of the experiment is restoring the image, a neural network has been used and trained to do this task. It takes 300x300 RGB image as an input and outputs an image with same structure.

Restoration CAE that has been used in experiments is modified/simplified version of the model proposed by Mao *et al.* (2016). It has been implemented using Keras framework in Python. It consists of 11 2D convolution layers, 2 max-poolings and 2 up-samplings. It has 1.814.628 trainable and 0 non-trainable parameters. In Figure 3.1 a visual representation of the model has been presented.

Model has been trained with images which have been distorted by all three types so that not only it can restore a singular distortion but also a group of them. Around 7500 frames have been used in training. These frames have been divided into four sets: original train, original test, distorted train and distorted test. Approximately for every test frame whether it's in original or distorted set there are two train frames in its counterpart. A distorted frame has been created using an original frame and distortion filters. So, they have been their exact counterparts. The numbers of frames in both train sets are equal so are the frames in test sets. Neural network first trained on original train and distorted train sets then validated the weights using test sets. Training has been done on GPU as time required for CPU would be too much. Training approximately lasted eight hours using a computer which has a GeForce GTX1080 graphical card.

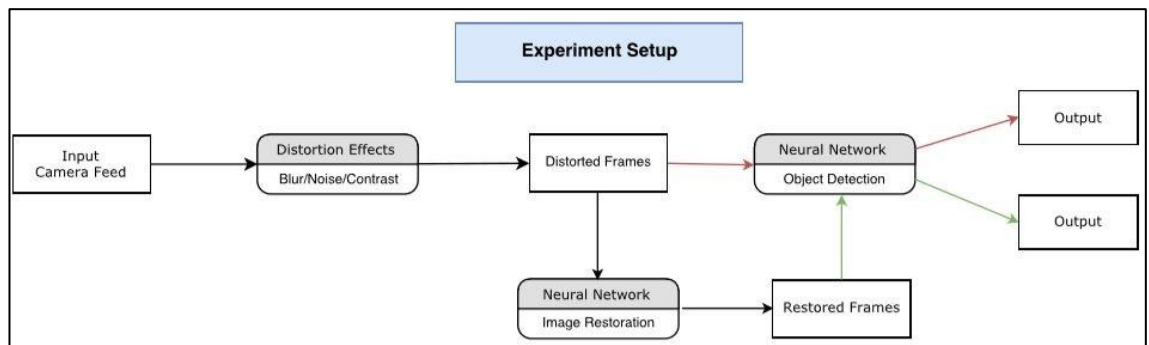
Figure 3.1: Restoration auto-encoder model architecture



3.5 METHOD

Only an iPhone 6 device has been used to gather the data, to distort frames and to utilize deep neural networks to limit the scope to mobile devices. In total around 360k frames from 20 different objects from various categories have been tested in showing distortion effects and 3 of them have been used in restoration stage. Overlook of the experiment setup has been given in Figure 3.2 below.

Figure 3.2: Experiment setup



3.5.1 Finding the Impact of Distortion Effects

First part of the experiment is about learning how the distortion filters effect CNNs. This part has 3 stages: preparation stage, distortion stage and comparison stage.

In preparation stage, utilizing each of four CNNs, object's reference values have been noted. For example, while classifying a banana, all four CNNs give the classification result with different confidence score. It is important to understand that confidence score is not the same for every CNN.

In distortion stage using OpenGL, distortion filters have been applied directly to camera feed. For blur, noise and contrast, 10 different distortion intensity levels have been varied in the tests.

After distorting the frames, from different angles, augmented frames have been passed to CNNs to classify the pictures again. For each distortion filter and its intensity level, an average confidence score of classification results from all four CNNs have been noted.

Lastly, in comparison stage difference of average confidence scores between distorted and original frame have been noted for all CNNs. These values also have been used to find the performance of restoration part.

3.5.2 Restoration Stage

In restoration stage, distorted frames of each distortion filters been attempted to restored using the CAE for picture restoration. Restored frames have been passed to picture classification CNNs and average confidence scores have been recorded.

The effects of all three distortion types have been tried to be rectified for all four deep network architectures. In total twelve experiment has been conducted using the three objects of the original twenty.

While testing for a specific distortion, two constant values have been given to other two distortions. These constants have been changed while testing for each distortion types. For example, while restoring a blurred image with different intensity, noise and contrast distortions have been given constant values to mimic a picture taken by average camera. These values have been found by comparing pictures taken by an average camera and pictures augmented by distortions. Figure 3.3 shows a sample of pictures taken by an average camera (General Mobile 5 Plus) and a quality camera (iPhone 6) but distorted by distortion filters to mimic the constant distortions in average cameras.

Figure 3.3: Raw and distorted pictures from average and quality cameras



The difference of average confidence scores between distorted and restored frame have been noted to find the performance of restoration CAE. Using the reference values given by original undistorted frames, distortion filters impact on classification confidence score and restoration CAE's compensation has been clearly shown.

Peak signal-to-noise ratio (PSNR) values have been given in restoring all three quality distortions. PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the quality of its representation. PSNR is generally used to measure the quality of reconstruction of signal and image. High PSNR usually means good image quality and less error is introduced to the image (Huynh-Thu and Ghanbari 2008).

The PSNR (in dB) is defined as:

$$\text{PSNR} = 10 * \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) \quad (3.1)$$

In this equation MAX_I is the maximum possible pixel value of the image. This value is 255 when the pixels in the image are represented using 8 bits per sample. Other term MSE is the mean squared error that is the difference between the predictor and what is predicted. For every distortion level in specific quality distortion two PSNR values have been presented by comparing original to distorted and original to restored frames.

4. FINDINGS

4.1 QUALITY DISTORTION EFFECTS

Most of the time image classification CNN's gives an array of object labels along with their confidence points. For example, when a banana is used as the input, CNN's output would be: 90 percentage banana, 9 percentage ruler, 1 percentage sponge.

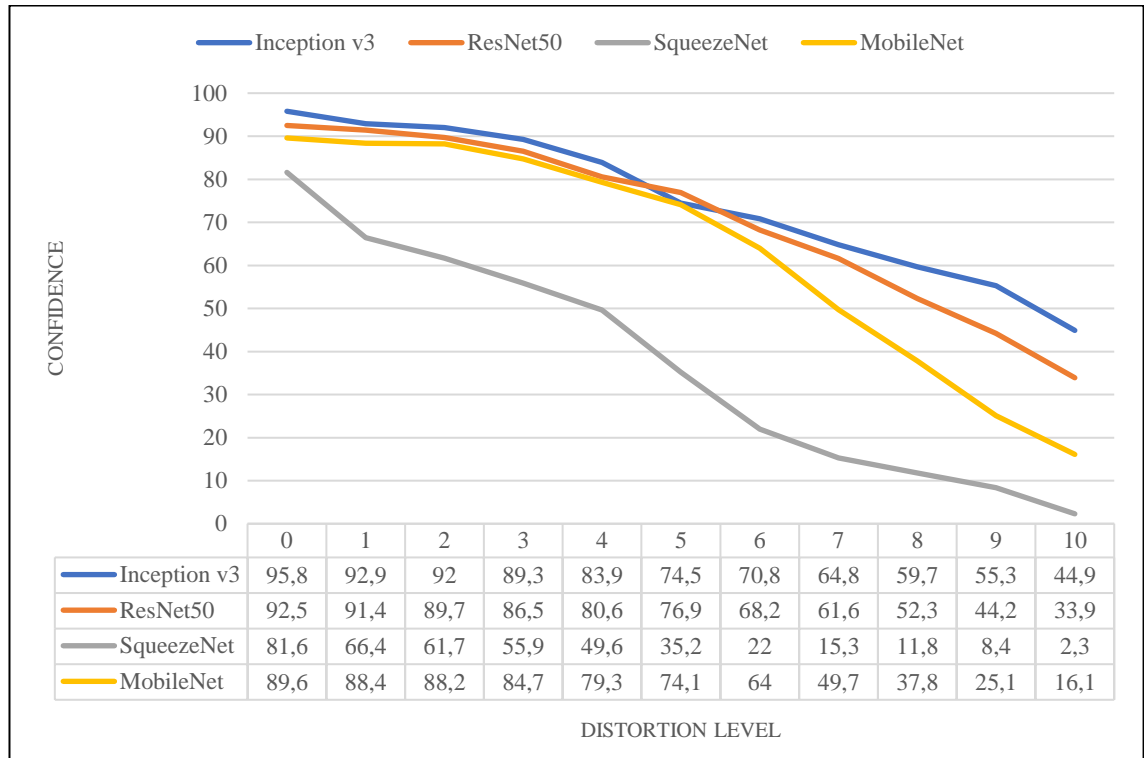
All twenty objects' confidence score values when introduced to quality distortions have given as impact tables in the Appendix section. The average calculated for all CNN architectures by summing the confidence scores for all levels and types of distortions separately then dividing them to 20, for SqueezeNet architecture it was divided to 18 as SqueezeNet has failed to detect the two objects in the dataset entirely.

In the figures below, average correct classification confidence point with distortion level is given for blur, noise and contrast. An effort has been made to ensure that during the experiment environmental conditions such as light intensity, angles and the distance between the camera and the object have been kept the same.

4.1.1 Blur

Figure 4.1 shows the change causing by the blur augmentation in frames. As seen in the figure, confidence point hasn't changed the same with each distortion level. When level 1 blur distortion was introduced for Inception v3 model, 2,9 percent change has been recorded. Then despite that level has been doubled only 0,9 percent change was observed. But each time level was increased especially after level 4, new wrong labels were started to be given in output array while the correct classification's confidence point decreased dramatically.

Figure 4.1: Blur distortion effects

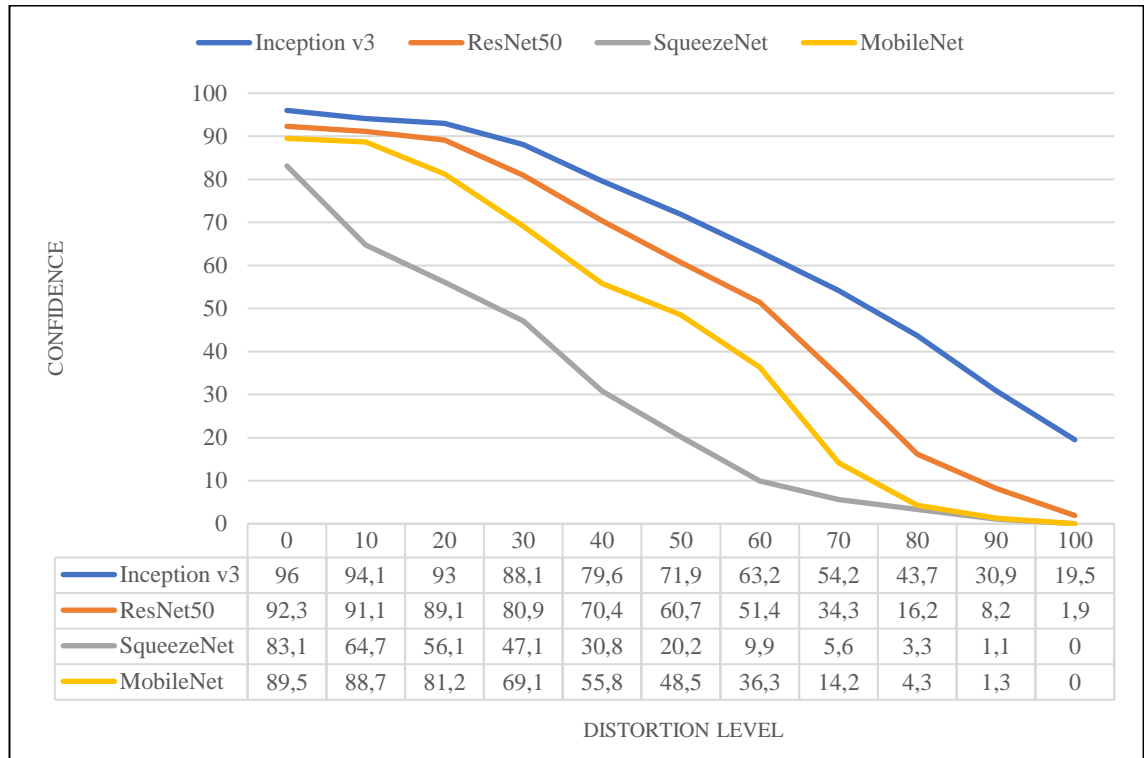


4.1.2 Noise

Next, noise distortion has been tested. As is seen in Figure 4.2, all of the networks are very sensitive to noise. Even though until level 20 most of the CNNs that have been tested has been somewhat resistant against distortion’s effects. But after 20, a dramatic decrease in confidence score has been observed.

However, compared with the SqueezeNet and MobileNet, it has noted that the performance of the Inception v3 and ResNet50 falls off slower. At a noise level of 80 the networks’ confidence scores have gotten to be less than 20% on average; despite this level of noise, the pictures were still effectively recognizable by human observers.

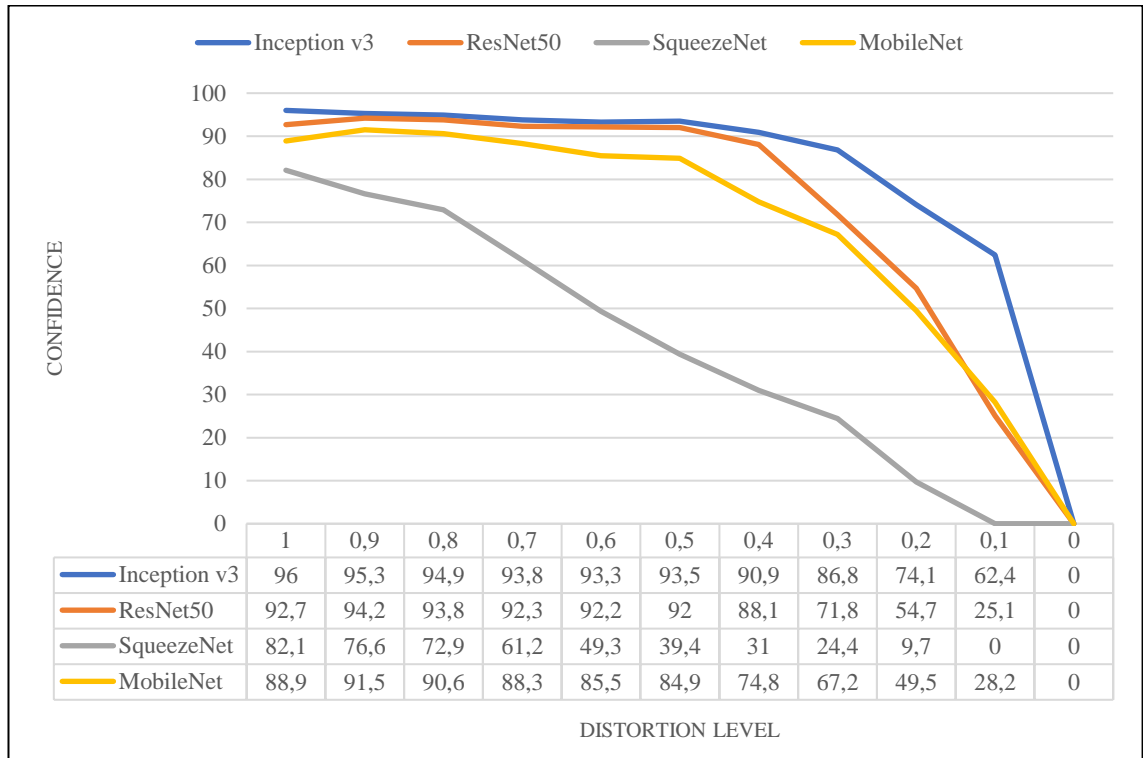
Figure 4.2: Noise distortion effects



4.1.3 Contrast

Lastly, effect of contrast changes on confidence score has been tested. As seen in Figure 4.3 except SqueezeNet, networks have been resilient with regard to contrast changes. Also, under high contrast distortion levels, recognizing pictures was as hard as to human observers as it was to CNNs.

Figure 4.3: Contrast distortion effects



It is observed that shape, size, detail level and how colorful is the object also effect confidence points against distortions. If the object is small like a band-aid or too detailed like a watch confidence points rapidly decreased every time level was increased. But if its big and not detailed like a plastic-bag or a white napkin, confidence points stayed the same for higher levels of distortion. Color is also a factor; it has been observed that the confidence points of objects like a bath-towel and a remote-controller with colorful buttons does not decrease the same against the distortions. CNNs quickly confused when a colorful object’s frames augmented by distortions given as input.

4.2 RESTORATION

For all four CNN architectures the effects of all three distortion types has been tried to be corrected. These experiments have been conducted in same environment where light condition is the same for all architectures. Also, care was taken to use similar angles and distance between camera and object. Even though restored pictures look all the same for

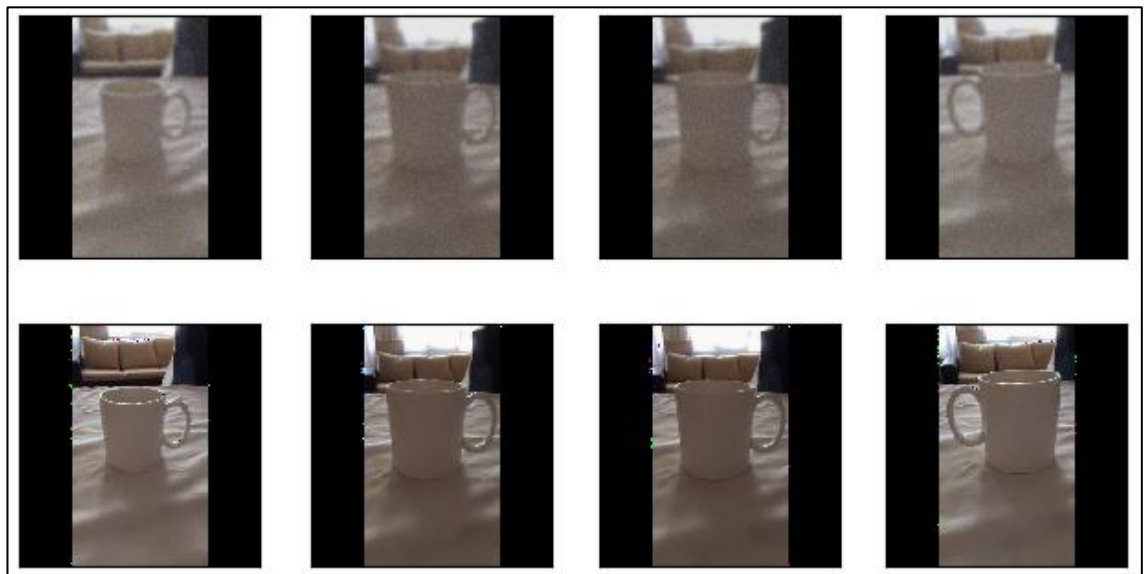
the human observer that's not the case for CNNs. Results have indicated that restored pictures perform different on each CNN and distortion type.

4.2.1 Blur

Restoring blurred images has been tested first for all four neural network architectures. "coffee mug" object is used in these experiments. Blur levels from 1 to 10 has been used while noise distortion level has been set to 20 and contrast to 0.8.

A sample of blurred and restored "coffee mug" images by restoration model is given in Figure 4.4. On sample image not only, blur but also other two distortions has been utilized.

Figure 4.4: Distorted and restored coffee mug



In Figure 4.5 PSNR values between original, distorted and restored frames for blur distortion has been given. Even though restoration in frames can be observed with naked eye, PSNR values between frames has been almost the same even though a small increase has been recorded.

Figure 4.5: PSNR values in restoration of blur distortion

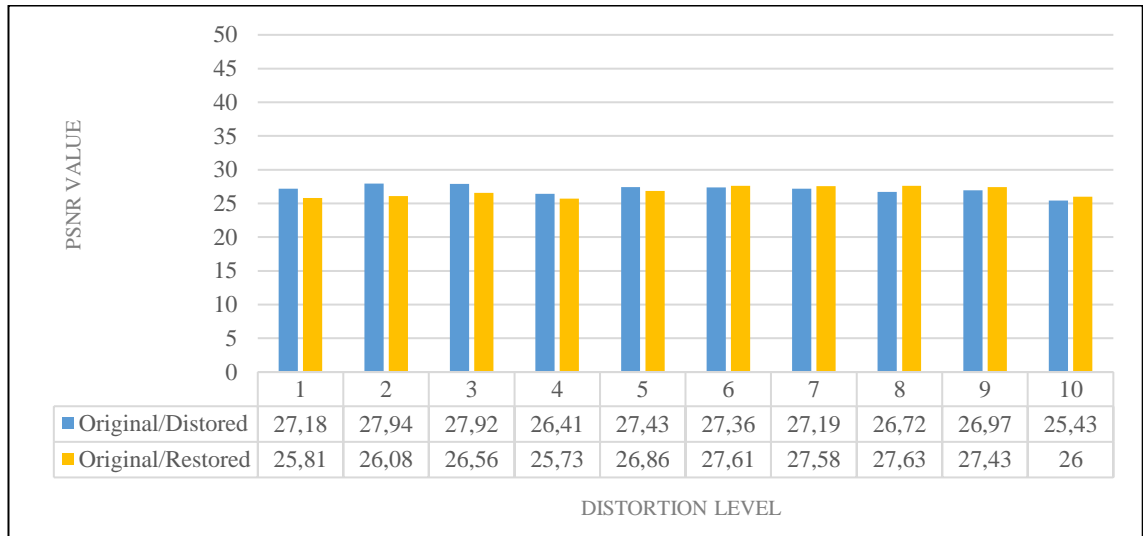
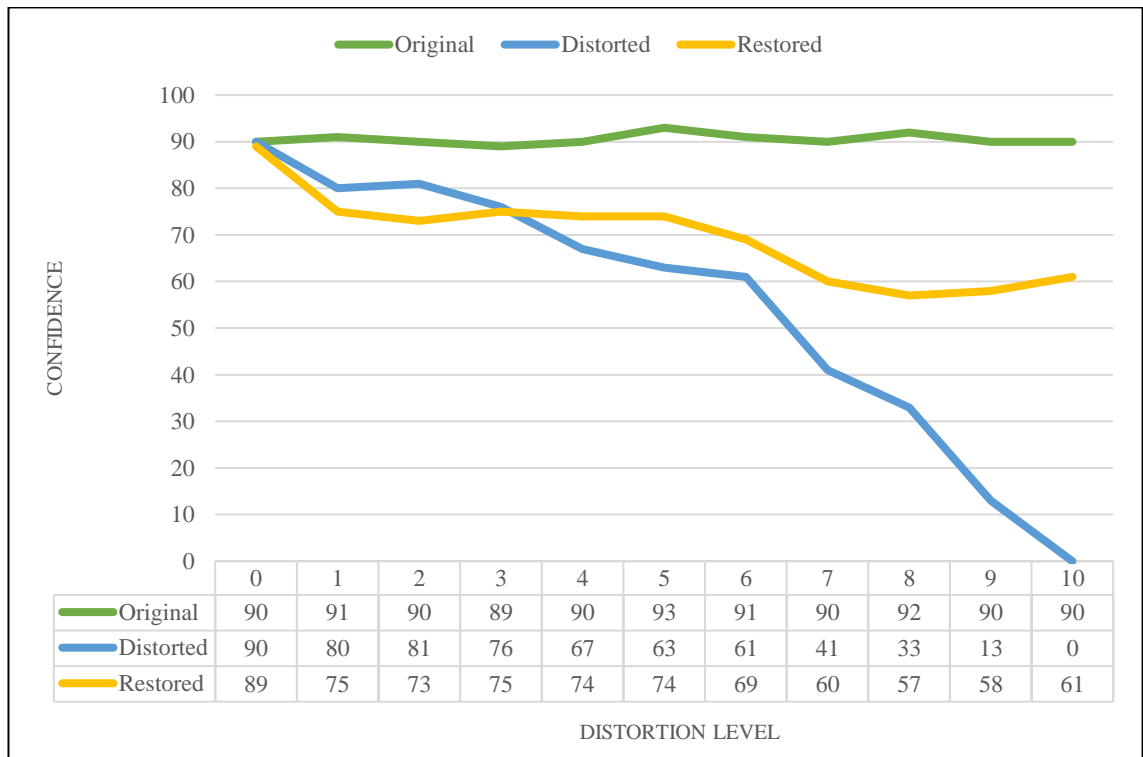


Figure 4.6 shows the confidence scores of original, distorted and restored camera frames of “coffee mug” object given by Inception v3 neural network architecture. As for all experiments, only the frames coming from real-time video feed have been used.

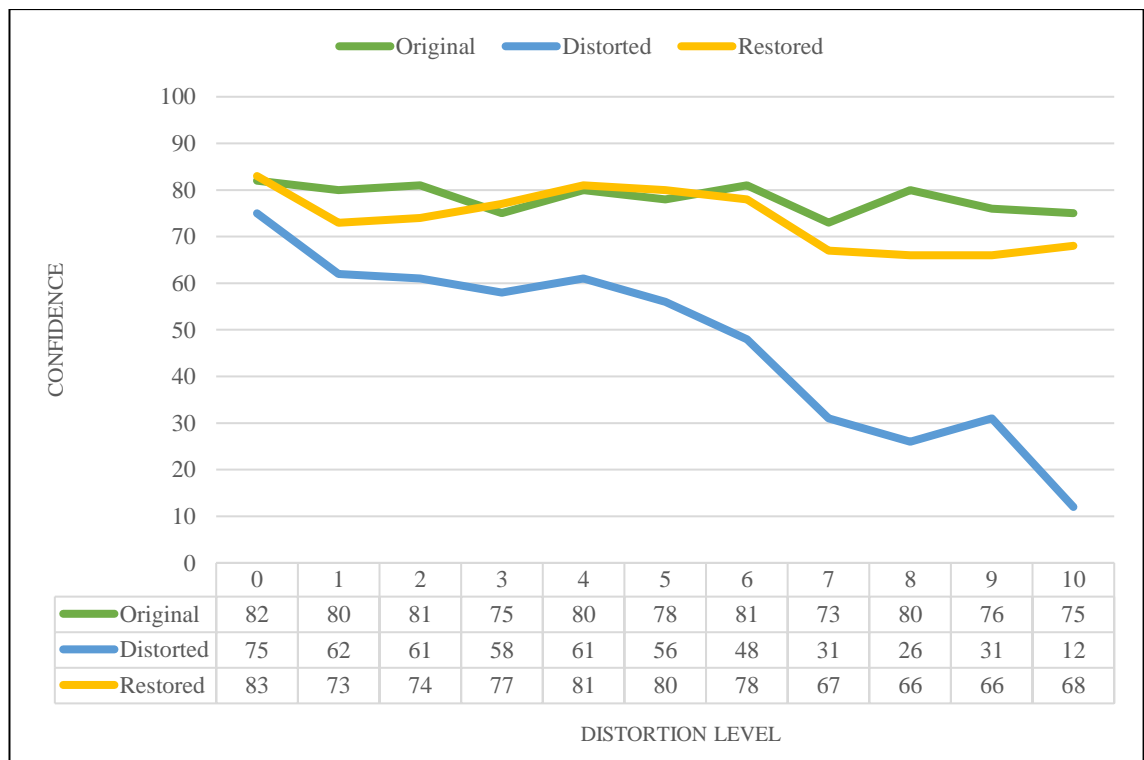
Figure 4.6: Blur restoration performance for Inception v3



It has been seen that restoration model couldn't correct the distorted frame 100% for this architecture. Moreover, it behaved like a distortion filter for early levels by dropping the confidence score. Be that it may, restoration model's confidence score stayed between 75 and 61. Even under maximum blur level where image classification no longer gives correct result for distorted frames, restoration model managed to stay in these bounds. Although it has seen that model hadn't give stable results when used with Inception v3. Especially after blur level 5 a significant drop has been seen in confidence score given by restoration model.

Confidence scores of original, distorted and restored camera frames of "coffee mug" object is given in Figure 4.7. It shows scores given by ResNet50 neural network architecture.

Figure 4.7: Blur restoration performance for ResNet50

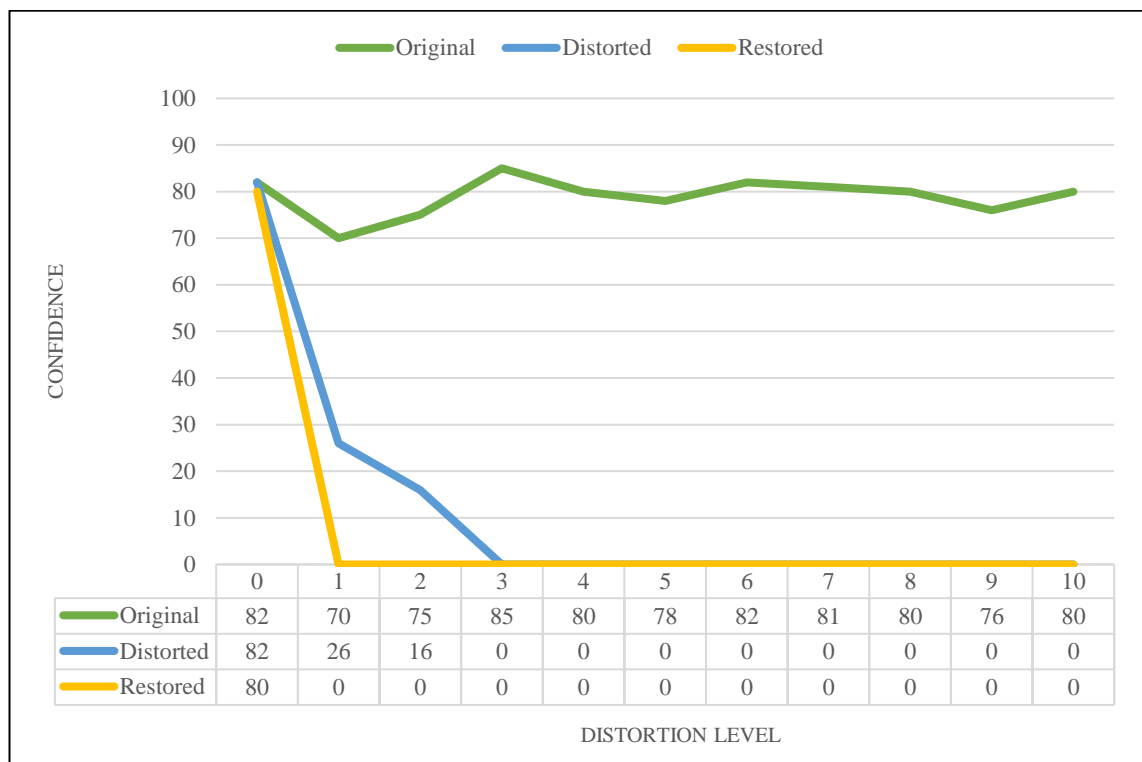


Similar to Inception v3 case, for early levels of blur distortion levels, restoration model has made the confidence score drop. Nevertheless, results indicate that restoration model corrects the effects of blur distortion. Even so, it has started to correct distorted frames from the beginning. Even at minor distortion such as level 1, a 10-point increase in

confidence score has been seen. Also, between distortion levels of 2 and 6 it has given almost same scores as original frames. In other words, restoration model fully corrected distorted frames for ResNet50 architecture for these levels. After level 7 blur distortion, a drop worth of 10 confidence points has been observed. But despite that this drop, model has given stable results for higher distortion levels.

Figure 4.8 shows the performance of restoration model in blur correction for SqueezeNet architecture. Same with others “coffee mug” object and real-time video feed has been used.

Figure 4.8: Blur restoration performance for SqueezeNet

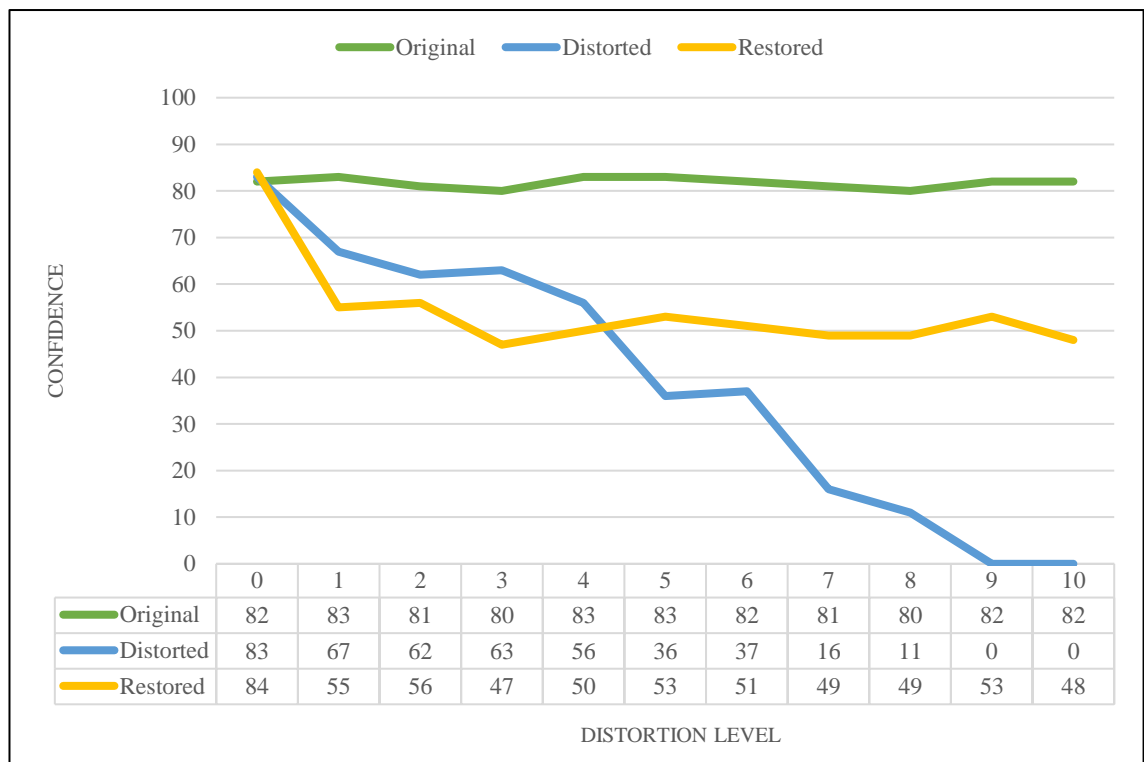


SqueezeNet is the most susceptible deep neural network architecture to image distortions. Even a small distortion is enough to drop significant confidence points. It also gave the worst performance with blur restoration model. Similar to other architectures restoration model has made the confidence points drop. But because SqueezeNet is more susceptible to distortions, restoration model made the “coffee mug” object unrecognizable. Despite that, it was recognizable for blur distortion level 1 and 2 though with low confidence

points. Results show that the restoration model has negative effect regards to confidence points in image classification and acts like a distortion effect.

Results of blur restoration model for MobileNet has been given at Figure 4.9. Same with other experiments with architectures distortion effect has been given in real-time and with those frames restoration model has been utilized.

Figure 4.9: Blur restoration performance for MobileNet



Restoration CAE model could not correct the blur distortions fully for MobileNet. Even though it had less impact on confidence points than blur distortions, it still caused some loss for earlier distortions.

Restoration model reached a break-through on blur distortion level 5 it. Increase in distortion level after 5 didn't have any effect on confidence points as model keep giving higher points than points caused by blur distortion. Moreover, though out the experiment restoration model gave steady results between 55 and 47 points. Still MobileNet's performance wasn't as good as Inception v3's and ResNet50's.

4.2.2 Noise

“Remote controller” object has been used in noise distortion restoration experiments. Noise levels from 10 to 100 has been utilized while blur distortion level has been set to 2 and contrast to 0.7 as constants.

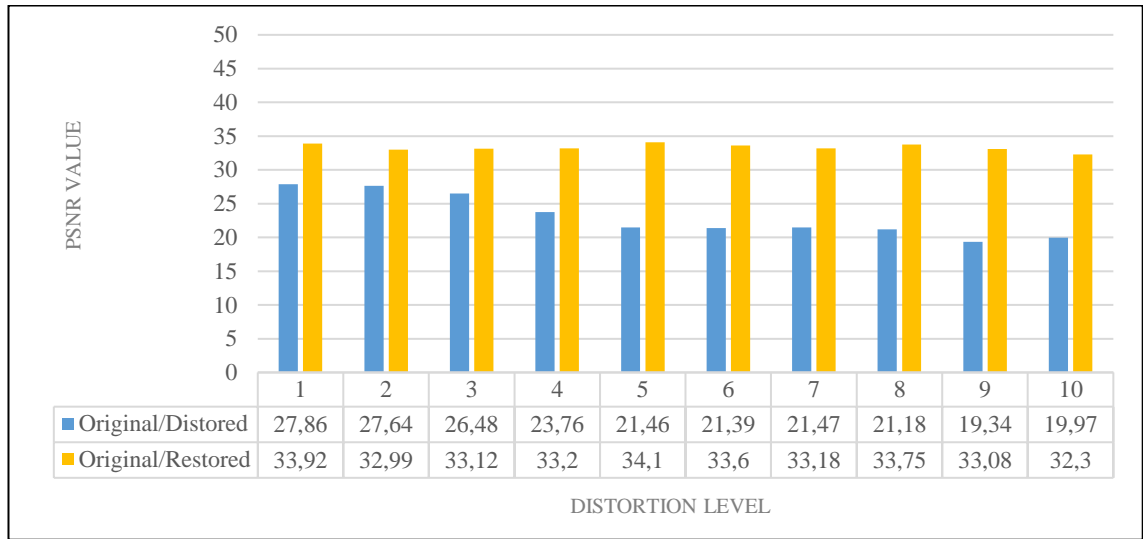
Figure 4.10 shows a sample of distorted and restored “remote controller” images. First row shows the distorted images by distortion filters such as blur, noise and contrast while the second row shows the restored images by restoration model.

Figure 4.10: Distorted and restored remote controller



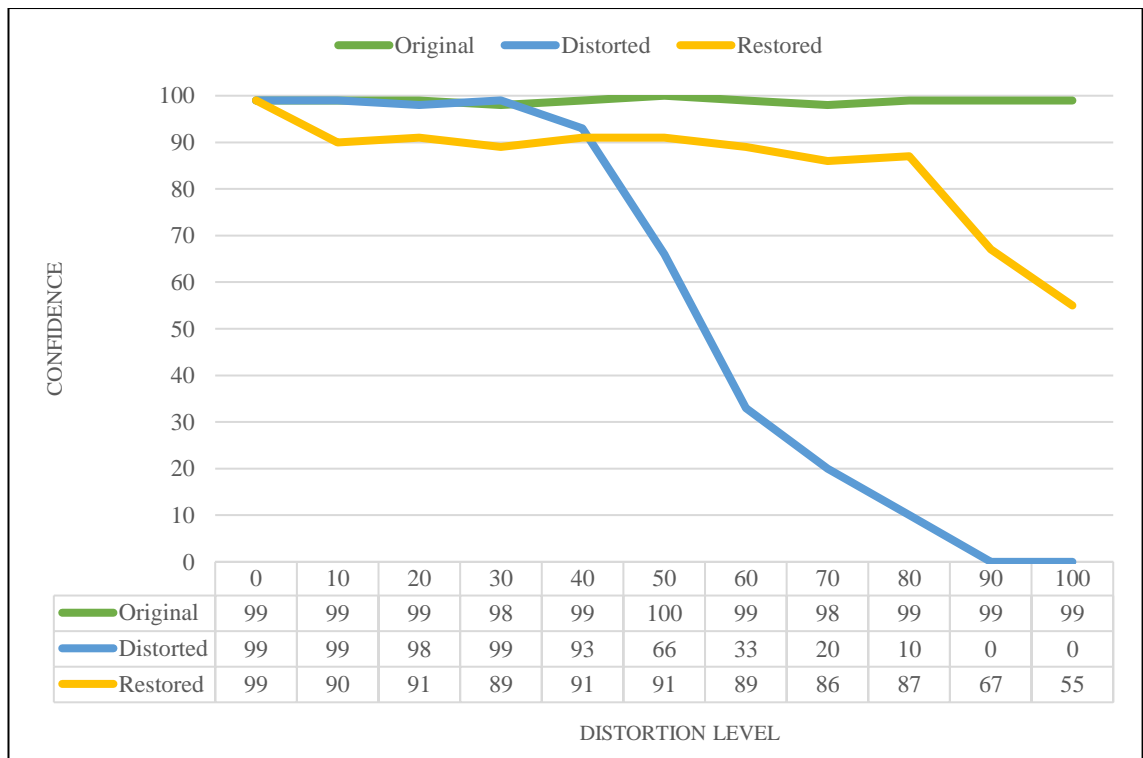
Figure 4.11 show the PSNR values between original, distorted and restored frames for noise restoration. From the first level of distortion a significant increase in PSNR has been recorded. As distortion levels increased, PSNR values compared original to distorted frames continuously decreased while values of original to restored frames have stayed above 32 dB.

Figure 4.11: PSNR values in restoration of noise distortion



Restoration model’s performance on Inception v3 architecture has been tested using distorted real-time camera frames of “remote controller” object. Figure 4.12 shows the impact of the restoration model on confidence scores of frames augmented by noise and other two constant distortions.

Figure 4.12: Noise restoration performance for Inception v3

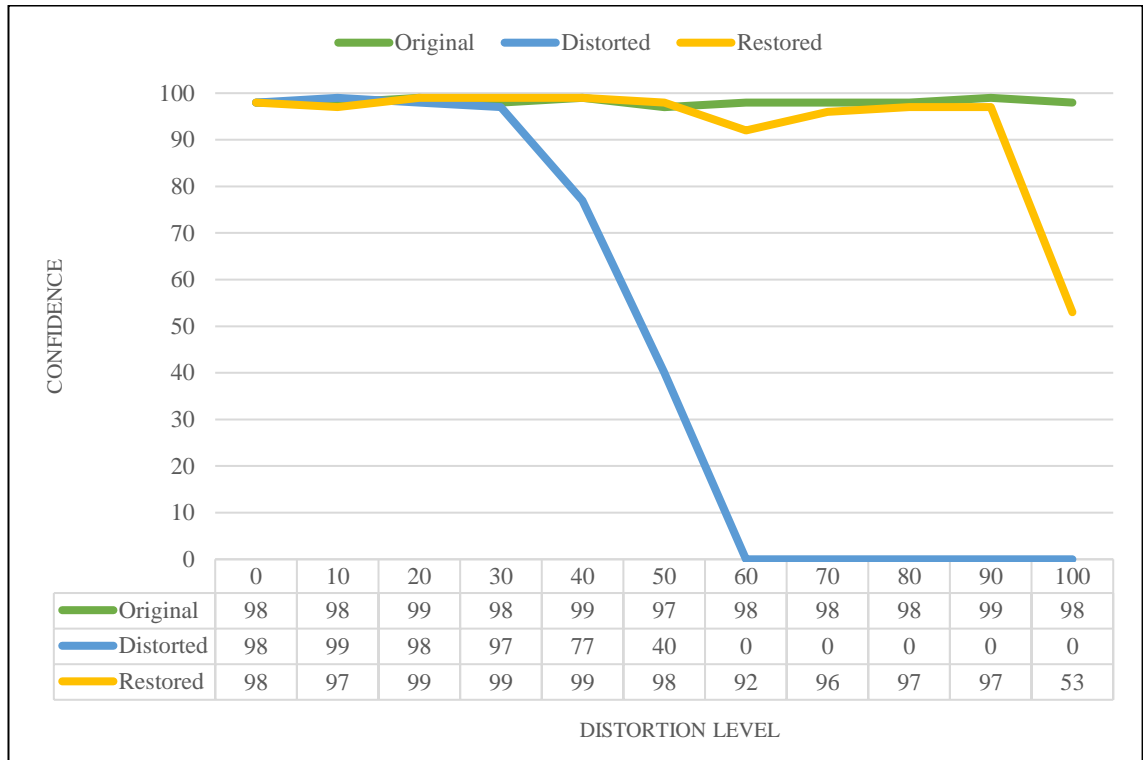


Confidence score of “remote controller” object has been very high. During the experiments it hasn’t go down below 98 for undistorted frames. In fact, the basic understanding of the Inception v3 model on this object is so deep that noise distortion has had little to no effect on confidence points up until level 50. Whereas for earlier distortion levels restoration model made the confidence score to drop up to 10 points. After noise level 50 where the image classification model based on Inception v3 architecture started to lose confidence points a break-through has been reached. Until distortion level 90, restoration model has given steady results almost the same with results has been given by original frames.

Nonetheless, a dramatic drop on confidence points has been observed after this level. Even so, for noise distortion levels 90 and 100 where model has been unable to recognize this object, restored images given by restoration model hasn’t go below 55 confidence score.

“Remote controller” object’s confidence scores given by its original, distorted and restored camera frames have been presented in Figure 4.13. It shows scores given by model based on ResNet50 neural network architecture.

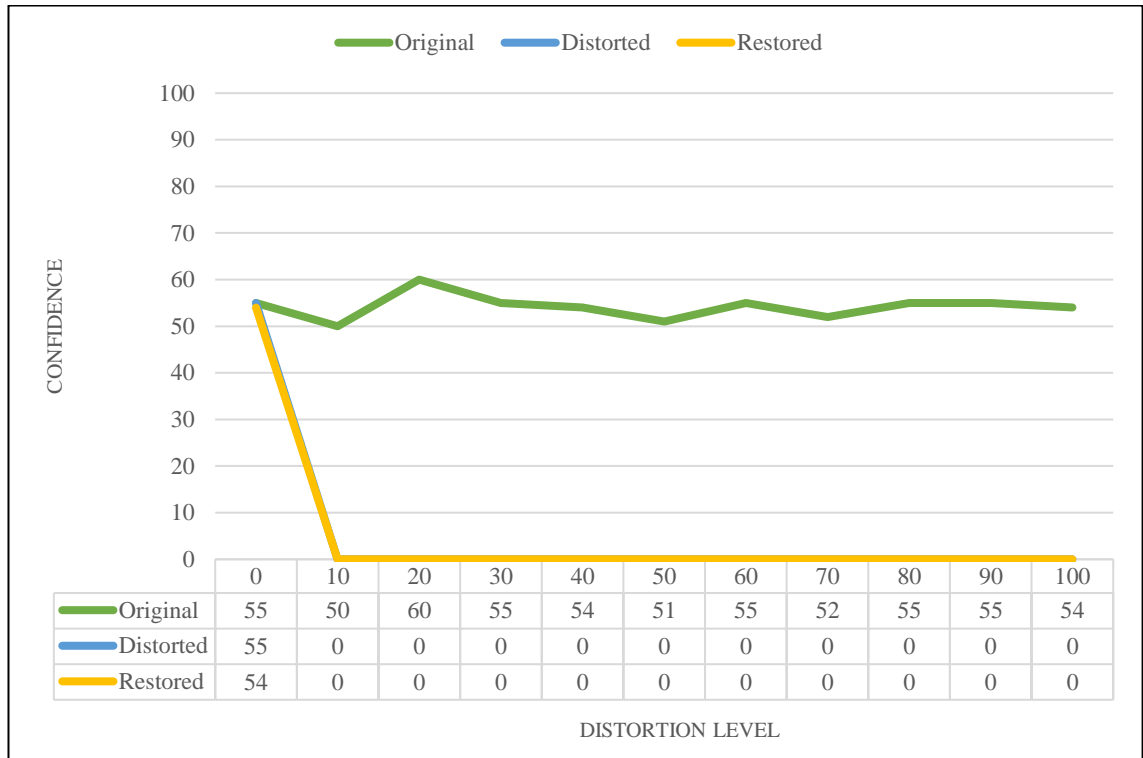
Figure 4.13: Noise restoration performance for ResNet50



Confidence scores of restored frames given by restoration model has been almost identical to original frames except for level 100 noise distortion. Unlike the results of other experiments restoration model has not caused a drop on confidence score for earlier distortions. Base understanding of the model which is based on ResNet50 architecture has been deep enough to resist the effects of noise distortion up until level 40. Results show that confidence score of restored frames given by image classification neural network model based on ResNet50 was the highest comparing to other three models.

In Figure 4.14 shows the restoration model’s performance on SqueezeNet architecture. Scores given by “remote controller” objects original, distorted and restored frames.

Figure 4.14: Noise restoration performance for SqueezeNet



As shown in first part of the paper, SqueezeNet performs worst under distortion filters. Beginning from the first level it couldn't identify the "remote controller" at all. As for the restoration model, it hasn't had any effect.

Lastly, MobileNet architecture-based model has been tested with restoration model. Like previous experiments care has been taken to keep same conditions, such as light, angles and distance from the camera. Figure 4.15 shows the results.

Figure 4.15: Noise restoration performance for MobileNet



Similar to ResNet50, base understanding of the MobileNet model on “remote controller” were deep as model’s confidence scores didn’t go below 98 points for original frames. Distortion augmentations didn’t have any effect on the model until level 30 noise distortion. From level 20 to level 30 a dramatic drop worthy of 60 points on confidence score has been observed.

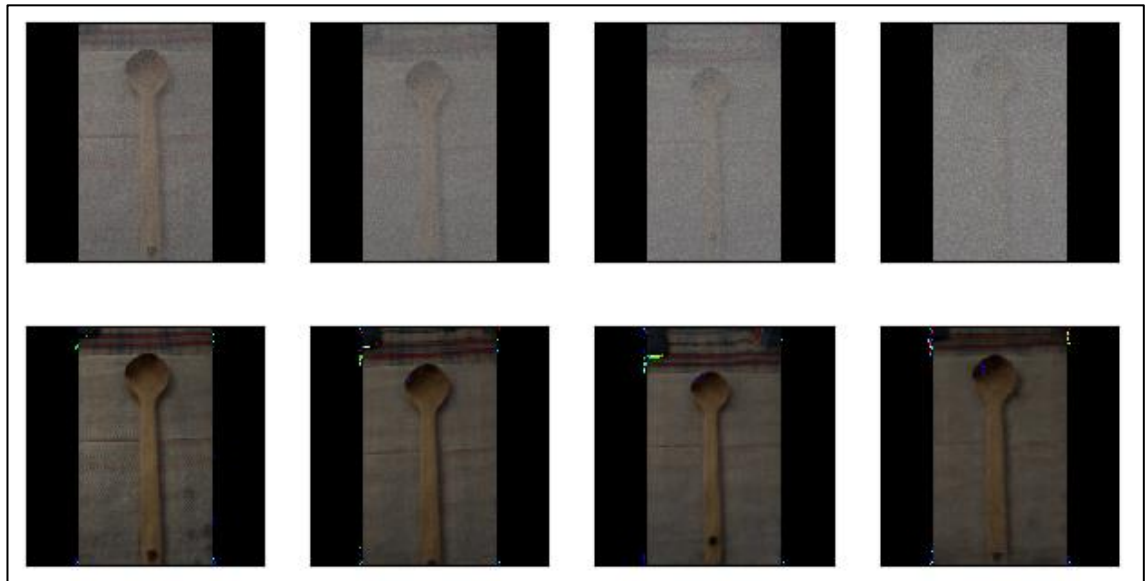
For MobileNet model, restored frames hasn’t caused a drop on confidence score while testing for earlier noise distortion levels. Moreover, restoration model’s restored frames have almost given same results as original ones. Up until the noise distortion level 70 results given by it were stable. Like the distorted frames’ confidence scores a notable drop worthy of 22 points has been observed when level 80 noise distortion introduced. On distortion level 100 confidence score dropped to 61 points.

4.2.3 Contrast

For contrast restoration experiments a “wooden spoon” has been used. Throughout the experiments environmental conditions were kept the same. So only the difference was the contrast distortion levels. Original, distorted and restored frames confidence scores have noted for all four neural network architectures. Starting from 1, contrast level of the frames gradually dropped to 0.1 while testing. Additional constant distortions also included, blur distortion has been set to 2 and noise to 30 for all frames.

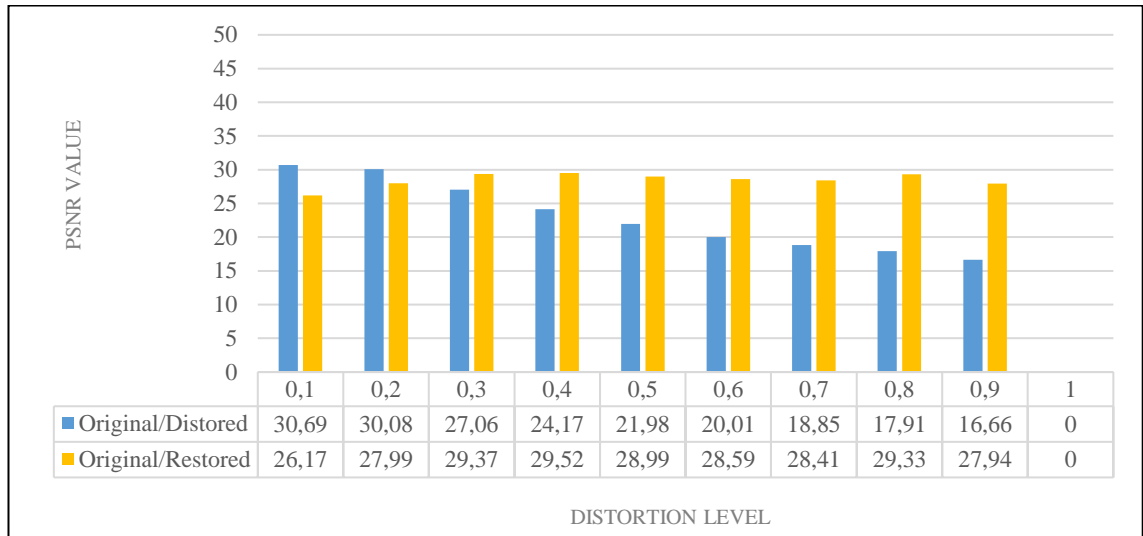
A sample of distorted and restored “wooden spoon” frames were given in Figure 4.16. On first row, distorted images by distortion filters such as blur, noise and contrast has presented while the second row shows the restored images by restoration model.

Figure 4.16: Distorted and restored wooden spoon



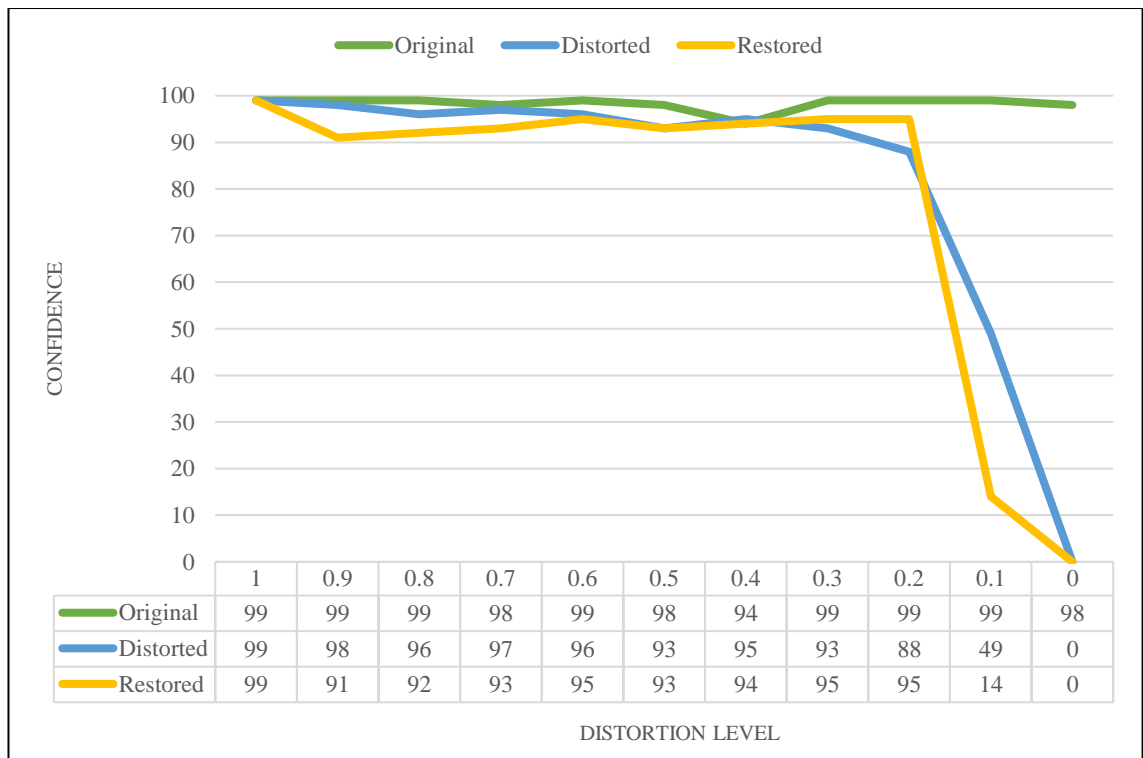
In figure 4.17 PSNR values for contrast distortion has been given. For first 3 levels original to distorted frames have given higher PSNR value which suggest restoration CAE distorted the frames more than distortion filter itself. Despite this situation, for higher levels of contrast distortion PSNR values of original to restored frames were greater.

Figure 4.17: PSNR values in restoration of contrast distortion



Restoration CAE model’s performance in terms of contrast correction on Inception v3 architecture has been tested using distorted real-time camera frames of “wooden spoon” object. Figure 4.18 shows the effectiveness of the restoration model on confidence scores on frames distorted by contrast and other two constant distortions.

Figure 4.18: Contrast restoration performance for Inception v3

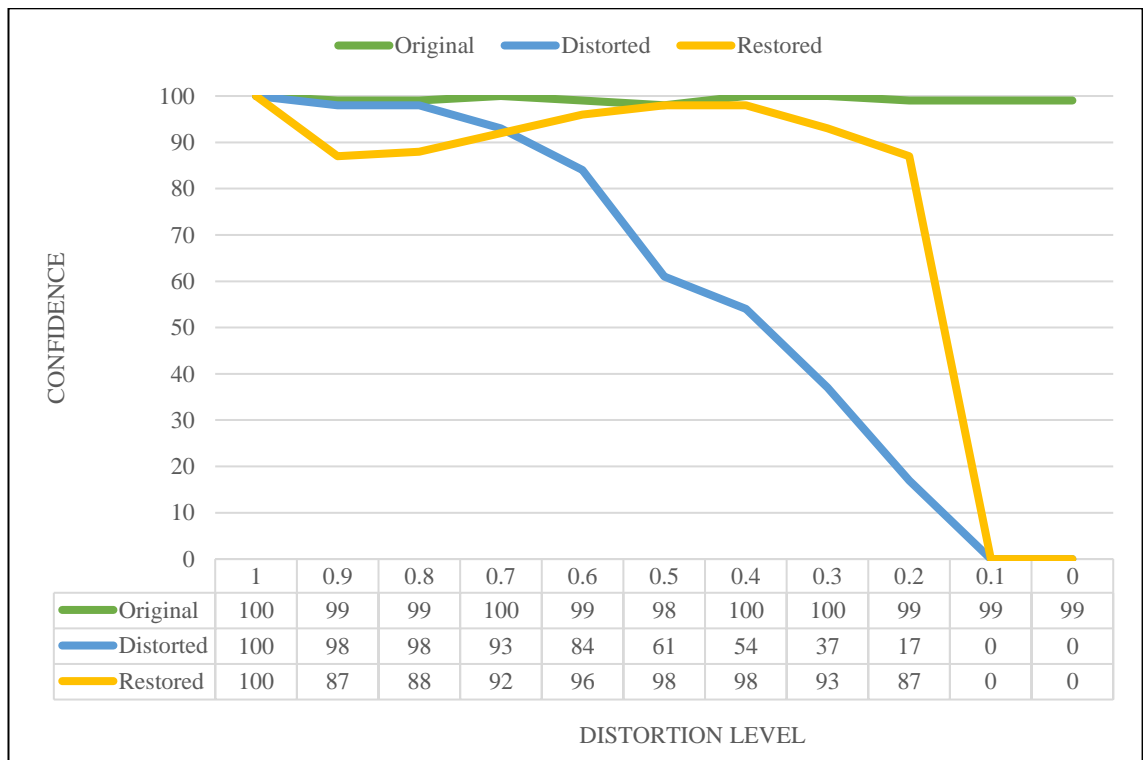


Restoration model managed give frames that increase the confidence score between distortion level 0.5 and 0.2. On distortion level 0.2, the increase in confidence score has been observed as 7 points.

A dramatic drop has been seen on confidence score when contrast of the frame had been set to 0.1. After this level, restoration model impacted the confidence score worse than contrast distortion itself.

Next, image classifier based on ResNet50 architecture model has been tested in contrast restoration. Original, distorted and restored “wooden spoon” object’s frames provided by real-time camera feed, distortion filters and restoration model. Results have been presented in Figure 4.19.

Figure 4.19: Contrast restoration performance for ResNet50

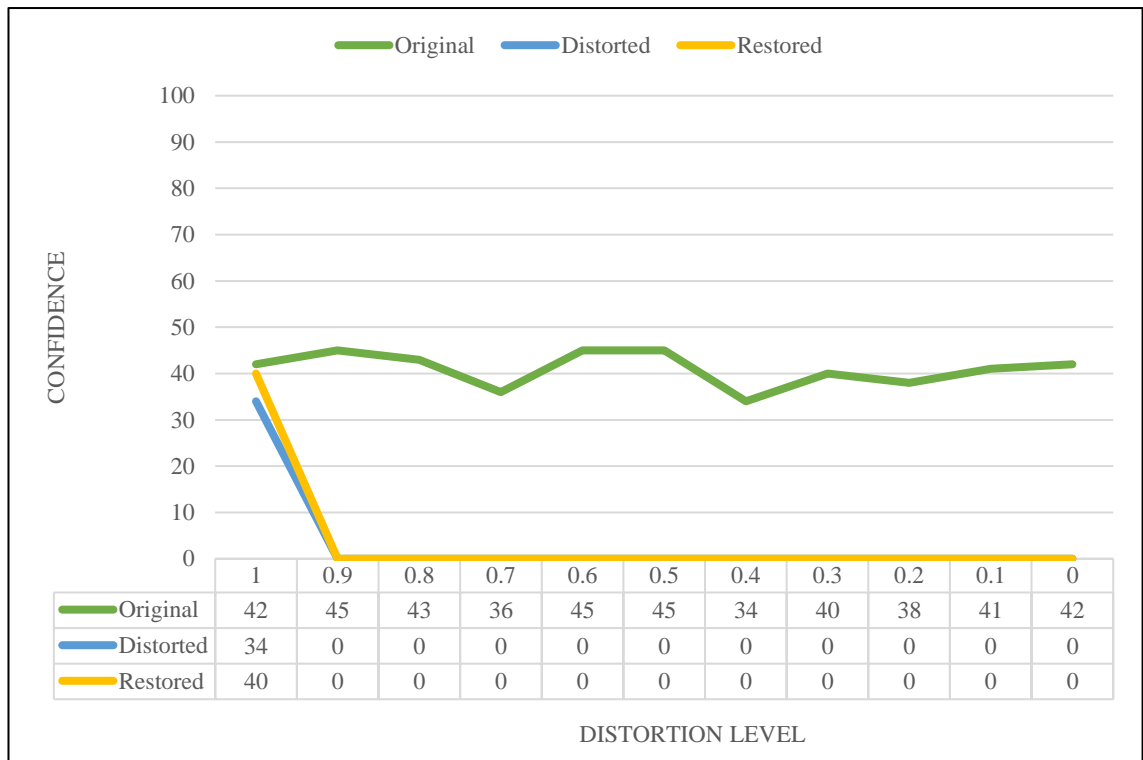


For earlier levels of contrast distortions, restoration model has cause a drop on confidence score. After the contrast has been set to 0.7, restoration model started to output frames that given same confidence score as distorted one. After this level of distortion, a

significant improvement worthy of 12 points has been observed. For higher levels of distortion until 0.1, restoration CAE almost completely corrected the contrast distortion.

Performance of restoration model for SqueezeNet has been tested for contrast correction with real-time camera frames of “wooden spoon”. Results of this experiment has been presented in Figure 4.20.

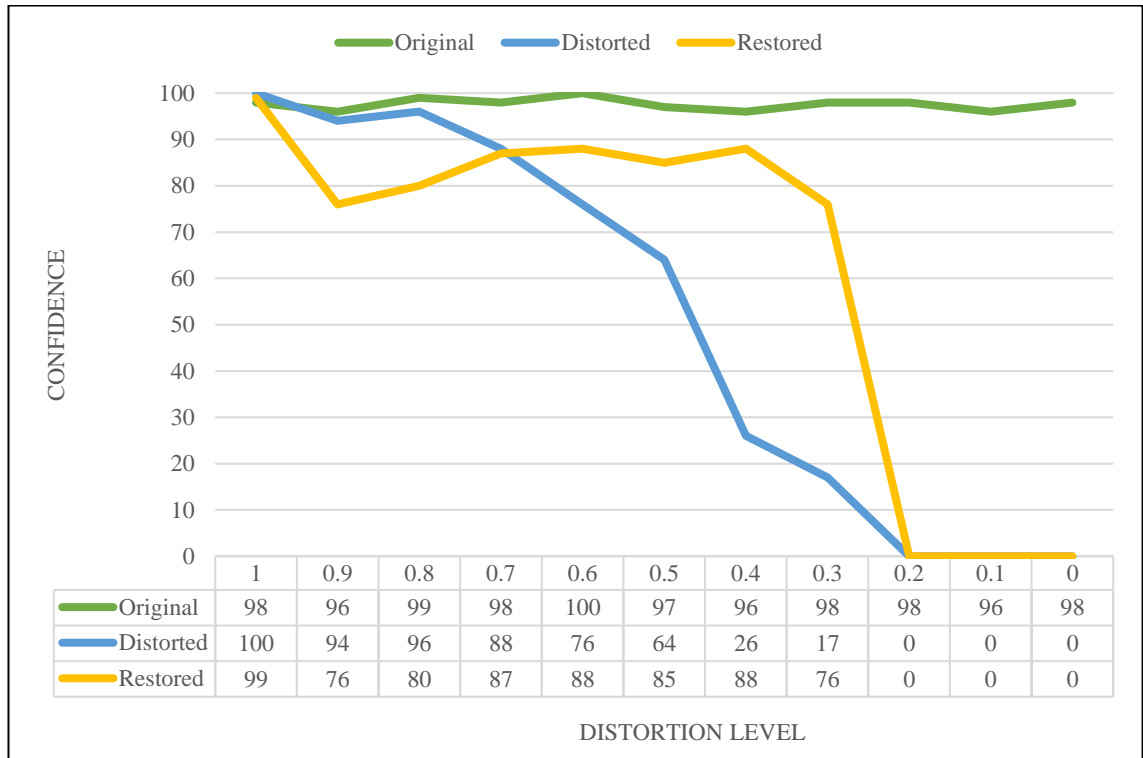
Figure 4.20: Contrast restoration performance for SqueezeNet



SqueezeNet model’s base understanding of this object was low from the beginning and it was so sensitive these distortions even from first level confidence score has dropped to 0. It has been found that frames provided by restoration model hadn’t had any effect on image classifier based on SqueezeNet architecture. Restored frames still were too distorted to increase the confidence score given by the image classifier CNN.

MobileNet architecture based model was tested last. Same with others original, distorted and restored frames of “wooden spoon” object has been used in contrast correction performance comparison. Figure 4.21 presents the result of this experiment.

Figure 4.21: Contrast restoration performance for MobileNet



Similar to ResNet50 architecture based image classifier’s case, restoration CAE had given frames which has performed worse than their distorted counterparts for early levels of distortion levels. On distortion level 0.7 restoration model has output frames that have same confidence score with distorted frames. After this level, restoration model has kept giving significantly better scores until contrast of the frame set to 0.2 then a dramatic drop has been observed and its frames also couldn’t be able to be recognized by MobileNet CNN.

Another finding about contrast restoration experiment is that when a background other than the trained one has been used confidence scores dropped from 90s to between 60 and 80s.

5. DISCUSSION

Results of experiments have been discussed in two parts. In the first part, impact of the distortion types and levels on four state-of-the-art CNN architectures for image classification has been reviewed. Suitability of the usage of restoration CAE, its effect on confidence score in terms of distortion and architecture of image classification have been assessed in the second part.

5.1 QUALITY DISTORTIONS

It has been found that distortion types, their levels and the image classification architecture all affect the confidence score. All four architectures lost confidence points when introduced distortion effects. This suggests that distortions affect the confidence score regardless of the architecture. But it has been found that some architectures show resilience to these effects and manage to keep their score by losing a few points. Also, it has been observed that distortion types, blur, noise and contrast don't disturb the results equally.

When introduced to images, blur causes a decrease in performance of CNNs by giving low confidence scores. In training CNNs learn features of the images, later they look for these to predict their class. Blur by default, damages the textures in the images, in these damaged textures, there could be features that are being looked for by CNN. This could be the reason of why blur affect the image classification CNNs. Although, for early levels of blur, except SqueezeNet all architectures showed some resilience. Being a relatively small model as it is only 5 MB, SqueezeNet's higher susceptibility to blur distortion effect could be explained by its lack of depth in the architecture. It is highly possible that depth of structure makes the CNN more resilient to blur effect. Because it learns more features from the images during the training. Even so, significant drops on confidence score have been recorded after level 5 and above, this fact suggests that depth of the architecture alone is not enough to resist impact of the blur effect.

Noise distortion effect has decreased the confidence score more than blur for all four CNNs. Inception v3 and ResNet50 showed some resilience until level 20 but after that they too started to lose confidence score rapidly with increasing levels. When advancing from distortion level 70 to 80 approximately 38% loss in confidence score has been observed while from 60 to 70 it was around 34%. From 90 to 100 it has become 78%. As a side note, even for level 100 noise distortion, frames were still clearly recognizable to human eyes. These values point that with each noise distortion level confidence score of correct class gets worse gradually. This situation could be indicating that not all features learned from the image are affected by noise equally. Like blur, it is possible that deeper structure of CNNs permits the networks to learn features that are resilient to noise for some levels. This understanding also explains why SqueezeNet is the most sensitive CNN to distortions as one of its benefits is reducing the depth of feature map to boost to computation.

As for the contrast distortion, except the SqueezeNet all other CNNs have managed to give high confidence scores for correct class labels. Even then contrast of the frames had been decreased by 70%, these other three CNNs managed to give approximately 76% confidence score. Which suggests features that are learned during the training are not easily distorted by contrast reduction if architecture's depth is normal not squeezed like SqueezeNet. It is also possible that because experiments have been done on real-time video stream, some features that are belong to specific angles and distances are more resilient to contrast reduction. In addition to that it has been observed shape, size, how detailed and colorful is object also a factor. If object is small or too detailed like a watch confidence score of correct class label rapidly decreased with every contrast distortion level. Also, it has been noted that colorful objects lose confidence score faster objects like bath-towel. But these levels of contrast distortions are not likely to be encountered on real-life scenarios. By testing couple of the low-tier devices, it has been seen that worst contrast distortion was around 0.7. At this level, average confidence score has been noted as 80.5%. What this value suggests is that it is possible to ignore contrast distortions in real-life usage.

From the experiments it has been observed that not all distortion types affect the CNNs equally, some CNNs have shown resilience for early levels of distortions, also different architectures give different results when encountered to those distortions which suggests it is possible to build picture classification CNN models that are more resilient to effects of quality distortions. Even though blur and noise still pose as a challenge to be solved because CNNs have been highly sensitive to them, results of contrast distortion can be ignored considering most CNNs were resilient to it.

5.2 RESTORATION CONVOLUTIONAL AUTO-ENCODER

Using a restoration CAE on mobile device to get better confidence score for correct class labels for the objects in real-time video stream has been tested in second part. Suitability of the restoration model for mobile device was a question in terms of function and performance. It has been shown that restoration CAE can be suitable for mobile usage and some increase in confidence score has been noted. But it was not the same for all architectures and distortion types and their levels. While testing the performance of the restoration CAE for specific distortion type, constant values have been set for other types to mimic real-life situation.

For blur restoration, except for SqueezeNet, increase in confidence score has been observed for all image classification architectures. Although for earlier levels of blur distortion, restoration CAE impacted confidence score worse than blur itself except ResNet50 architecture. It is possible that restoration model distorts these frames because of lack of training as experiment has been done on real-time video using mobile device. Restoration CAE takes a frame as an input and generates a new frame using features learned during training. It is likely restoration model couldn't generate every feature of the original frame so for early levels it acts like a distortion filter. For ResNet50 case, even though restoration model causes some distortion, it fully corrected the distorted frames which suggests image classifier CNN architecture also determines performance of the restoration. As for other two architectures, after early levels a break-through has been reached at different distortion levels. For Inception v3 it has been after level 2 while for MobileNet it has been after level 5 which indicates that even though restoration model

causes loss in confidence score for early levels its possible that at some level it may restore more features than distortion filter damages.

Noise restoration performance of restoration CAE model has been better than blur restoration performance. For higher levels of noise distortions while image classifier CNNs were no longer identify the correct class label, restoration model managed to give over 90% confidence score for some models. Best results have been observed when using ResNet50 architecture based image classifier while the worst with SqueezeNet. Despite that they look the same for the human eye, restored frames may not have the features needed by CNN to classify, another explanation would be none of features of the object that SqueezeNet learned during the training couldn't be generated by the restoration model because of its depth does not allow it learn features that are more robust to some minor modifications. In noise restoration experiment only for Inception v3 it acted like a distortion filter for early levels for blur it was the same case for MobileNet too which suggests for different distortion types performance of the restoration model may change. It is certain that architecture of image classifier has a direct effect on restoration model's performance also for restoring noised images.

Results of the contrast restoration has been quite different from the blur and noise. Restoration model has given better scores for some distortion levels but mostly after level 0.7 before that it caused some drop on confidence score. For SqueezeNet no performance gain has been noted. It is possible that restoring color is different from restoring blur and noise distortions. Getting better confidence score for mid and high contrast distortion levels on Inception v3 and ResNet50 suggests that restoring contrast is possible for these levels. Although, unlike noise and blur distortions it is not sustainable for higher levels such as 0.2. Together with this and MobileNet's case once again shows that architecture of the CNN plays an important role on the performance of the restoration model. Also, it should be noted that all image classification CNNs showed resistance for early levels of contrast distortions and for higher levels restoration CAE performed better. So, to create more robust visual systems against quality distortions, it might better to use restoration CAEs where CNNs give lower confidence scores than CAEs.

Almost for all architectures, restoration CAE has impacted confidence score worse than distortion filter itself for early levels of distortions. It is possible that restoration model distorts these frames because of lack of training as the experiment has been done on real-time video using mobile device. Restoration CAE takes a frame as an input and generates a new frame using features learned during training. It is likely that restoration model could not generate every feature of the original frame so for early levels it acts like a distortion filter but for higher levels of distortion restoration CAE generates more features than distortion filter damages so increase in confidence score occurs. In SqueezeNet architecture 3x3 filters are replaced with 1x1 filters to reduce the number of parameters by 9x. Typically, a larger 3x3 convolution filter captures spatial information of pixels close to each other. On the other hand, 1x1 convolutional filters zero in on a single pixel and capture relationships amongst its channels as opposed to neighboring pixels. This could be reason that restoration CAE could not improve the confidence score of SqueezeNet for all three distortions.

Response time for a single frame of CAE has been around 7.6 seconds. Although this value would be considered high, it is possible to reduce the time needed by doing more optimization in convolution layer depth and filter size. Also using efficient neural network architectures especially designed for devices with limited computation power may significantly improve the performance in terms of response time.

In the beginning of the paper, finding quality distortions' impact on different image classifier architectures optimized for mobile usage on real-time video stream and whether their types would cause different performance outcome, learning the usability of the restoration CAE on mobile device and its performance on different architectures and quality distortions were the initial goals. These goals have been achieved with experiments and comparisons that have been done. Because this work covers the state-of-the-art CNN architectures optimized for mobile devices result of this research enables future work in developing machine vision systems on mobile devices that perform more unchanging to quality distortions and better restoration CAEs which can correct these distortions by taking account of image classifier architecture and distortion types and gives better insight about how quality distortions affect CNNs.

6. CONCLUSIONS AND FUTURE WORK

An evaluation of four state-of-the-art deep neural network models for picture classification under quality distortions on mobile environment has been made to understand whether different architectures would cause a difference in terms of confidence scores of correct class labels. Impact of these distortions on image classification results have been compared using different models.

In gathering the data, distorting the frames and utilizing deep neural networks, only an iPhone 6 device has been used to limit the scope for mobile devices. In total around 360k frames from 20 different objects from various categories have been tested in showing distortion effects and 3 of them have been used in restoration stage.

First part of the experiment has been about learning how the distortion filters effect CNNs. This part consisted of three stages: preparation stage, distortion stage and comparison stage.

In preparation stage, using different angles and distances object's reference values have been gathered utilizing each of four CNNs and average value has been noted.

In distortion stage distortion filters have been applied directly to camera feed using OpenGL. 10 different distortion intensity levels for blur, noise and contrast have been varied in the tests. After distorting the frames, from different angles, augmented frames have been passed to CNNs to classify the pictures again. For each distortion filter and its intensity level, an average confidence score of classification results from all four CNNs have been noted.

Lastly, in comparison stage difference of average confidence scores between distorted and original frame have been noted for all four CNNs. The impact of the quality distortions clearly shown with these values. These values also have been used to find the performance of restoration part.

It has been observed architecture of picture classification CNN directly effects the quality distortions' impact on confidence score. Different quality distortions affect the confidence score differently. The variance of objects that have been used in experiments suggests that it is independent from the object used. Based on the results from different CNNs that have been used in these experiments, it is possible to create CNN architectures that are more invariant to quality distortions. This would be possible to create models that learn the features in a way that they are not easily affected when encountered to quality distortions such as blur, noise and contrast.

As second part, suitability of using restoration CAE on mobile devices have been investigated. Then classification results of restored images have been assessed in terms of image classification architectures, distortion types and levels. In this part, distorted frames of each distortion filters been attempted to restored using the CAE for picture restoration. Restored frames have been passed to picture classification CNNs and average confidence scores have been recorded. The effects of all three distortion types have been tried to be rectified for all four deep neural network architectures. In total twelve experiments have been conducted using the three objects of the original twenty.

While testing for a specific distortion, two constant values have been given to other two distortions. These constants have been changed while testing for each distortion types. These values have been found by comparing pictures taken by average a camera and quality pictures augmented by distortion filters.

It has been seen that in mobile device, it is possible to use restoration CAE for real-time video stream's distorted frames to improve confidence score of correct class label. With model used in these experiments, restoring a single frame took approximately 7.6 seconds. Although this response time of restoration CAE might be considered high, creating restoration models that are optimized for mobile usage is achievable like it has been for image classification CNNs.

Like the impact of the quality distortions, architecture of image classification CNN directly affects the restoration CAE performance. Even though restored frames look all the same for human observers, features that are required by some image classification

CNNs may not be restored correctly or not restored at all. Using architectures that are compatible with the image classification CNN architecture would increase visual system's resistance against quality distortions. If the intention is to use this model in mobile environment researcher must keep in his/her mind that response time and resource management play a significant role as well on the performance of the CAE.

Also, it has been noted that CAE performs differently for each different quality distortions and their levels. While it significantly increased the confidence scores of distorted frames by blur and noise, it wasn't same case for contrast restoration. But almost all image classification CNNs showed resistance for early levels of contrast distortion and restoration CAE performed better where CNNs couldn't, so to create more robust visual systems against quality distortions, it might better to use restoration CAEs where image classification CNNs give lower confidence scores than CAEs.

To sum up, in mobile environment real-time video stream is likely to be experience quality distortions because of movement, weather conditions and poor-quality camera lens. Different deep neural network architectures give different results as some of them are more sensitive to the quality distortions while others manage to keep invariant for early and medium levels of distortions which suggests it is possible to build machine vision systems that are optimized enough to run on mobile devices by choosing or creating deep neural network architectures that are resistant against these distortions and with making sure that they are compatible with each other utilizing restoration CAE where they fail. In creating these models, especially if the intention is using these in mobile environment, optimization on resource usage and response time should be considered.

As for future work, developing an image classifier CNN architecture that gives more accurate results and resistant against quality distortions is considered. Furthermore, compatible to this CNN, creating a faster and more precise deep convolutional auto-encoder restoration model that is optimized for mobile usage is another objective for future work.

REFERENCES

Periodicals

- Haerberli, P., & Voorhies D., 1994. Image Processing by Linear Interpolation and Extrapolation. *IRIS Universe Magazine*. **28**, pp. 8–9.
- Hinton, G. E., and Salakhutdinov, R. R., 2006. Reducing the Dimensionality of Data with Neural Networks. *Science*. **313** (5786), pp.504-7.
- Huynh-Thu, Q., Ghanbari, M. 2008. Scope of Validity of PSNR in Image/Video Quality Assessment. *Electronics Letters*. **44** (13), pp. 800–801.
- Ji, S., Xu, W., Yang, M. and Yu, K., 2013. 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **35** (1).
- LeCun, Y., Bengio, Y., & Hinton, G., 2015. Deep learning. *Nature International Journal of Science*. **521**, pp. 436-444.
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P., 1998. Gradient-based Learning Applied to Document Recognition. *Proceedings of the IEEE*. **86**(11), pp. 2278-2324
- Masci J., Meier U., Ciresan D., Schmidhuber J., 2011. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. *ICANN 2011. Lecture Notes in Computer Science* **6791**, pp.52–59.
- Rumelhart, D.E., Hinton, G.E., & Williams, R.J., 1986. Learning Representations by Back-Propagating Errors. *Nature*. **323**, pp. 533–536.

Other Publications

- Borkar, T. and Karam, L., 2017. DeepCorrect: Correcting DNN models against Image Distortions. *arXiv:1705.02406 [cs.CV]*.
- Ciresan, D.C., Meier, U., & Schmidhuber, J., 2012. Multi-column deep neural networks for image classification. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3642-3649.
- Collobert, R., and Jason W., 2008. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask. *ICML '08 Proceedings of the 25th international conference on Machine learning* pp. 160-167.
- Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Fei-Fei, L., 2009. ImageNet: A Large-Scale Hierarchical Image Database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248-255.
- Dodge, S.F., & Karam, L.J., 2017. A Study and Comparison of Human and Deep Learning Recognition Performance under Visual Distortions. *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, pp. 1-7.
- Fu, J., Wang, H., & Zuo, L., 2016. Blind Image Quality Assessment for Multiply Distorted Images Via Convolutional Neural Networks. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1075-1079.
- Gao, R., & Grauman, K., 2017. On-demand Learning for Deep Image Restoration. *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 1095-1104.
- Goodfellow, I.J., Shlens, J., & Szegedy, C., 2014. Explaining and Harnessing Adversarial Examples. *International Conference on Learning Representations (ICLR), 2015*.
- He, K., Zhang, X., Ren, S., & Sun, J., 2016. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778.
- Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H., 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv:1704.04861 [cs.CV]*.

- Iandola, F.N., Moskewicz, M.W., Ashraf, K., Han, S., Dally, W.J., & Keutzer, K., 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size. *arXiv:1602.07360 [cs.CV]*.
- ImageNet Official Website. 2010. [online] <http://image-net.org/about-stats>. [accessed on 08 June 2018].
- Krizhevsky, A. 2009. Learning Multiple Layers of Features from Tiny Images. *Thesis for MSc Degree*. Toronto: University of Toronto.
- Mao, X., Shen, C., & Yang, Y., 2016. Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections. *arXiv:1606.08921[cs.CV]*.
- Nguyen, A.M., Yosinski, J., & Clune, J., 2015. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 427-436.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A., 2015. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1-9.
- Zhang, X., Zhou, X., Lin, M., & Sun, J., 2017. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. *arXiv:1707.01083 [cs.CV]*.
- Zhou, Y., Song, S., & Cheung, N., 2017. On Classification of Distorted Images with Deep Convolutional Neural Networks. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1213-1217.

APPENDICES

APPENDIX A.1 Table 1, Wooden spoon - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	99%	99%	94%	100%
1	100%	100%	95%	100%
2	100%	100%	94%	100%
3	99%	100%	83%	100%
4	97%	96%	71%	98%
5	93%	88%	24%	91%
6	94%	82%	9%	73%
7	94%	65%	0%	49%
8	89%	48%	0%	17%
9	83%	20%	0%	0%
10	78%	0%	0%	0%

APPENDIX A.1 Table 2, Wooden spoon - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	100%	100%	95%	100%
10	99%	100%	90%	100%
20	100%	100%	84%	100%
30	100%	99%	84%	99%
40	99%	99%	73%	95%
50	99%	83%	37%	80%
60	99%	70%	0%	54%
70	93%	26%	0%	0%
80	80%	0%	0%	0%
90	50%	0%	0%	0%
100	31%	0%	0%	0%

APPENDIX A.1 Table 3, Wooden spoon - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	99%	100%	97%	100%
0.1	99%	100%	95%	100%
0.2	99%	100%	93%	100%
0.3	99%	100%	87%	99%
0.4	99%	99%	43%	99%
0.5	100%	98%	20%	99%
0.6	99%	95%	0%	99%
0.7	99%	69%	0%	88%
0.8	97%	38%	0%	36%
0.9	81%	0%	0%	0%
1.0	0%	0%	0%	0%

APPENDIX A.2 Table 1, Washbasin - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	99%	99%	94%	100%
1	99%	100%	96%	100%
2	99%	100%	94%	99%
3	96%	99%	83%	95%
4	96%	99%	50%	94%
5	93%	94%	37%	93%
6	93%	83%	13%	86%
7	87%	73%	0%	77%
8	88%	61%	0%	25%
9	83%	41%	0%	14%
10	78%	24%	0%	0%

APPENDIX A.2 Table 2, Washbasin - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	100%	100%	95%	100%
10	100%	100%	95%	100%
20	100%	99%	96%	100%
30	98%	100%	95%	99%
40	99%	99%	80%	95%
50	98%	97%	81%	90%
60	100%	98%	71%	80%
70	99%	80%	50%	52%
80	80%	53%	30%	33%
90	64%	42%	10%	13%
100	30%	19%	0%	0%

APPENDIX A.2 Table 3, Washbasin - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	99%	100%	97%	100%
0.1	100%	99%	98%	99%
0.2	99%	100%	99%	99%
0.3	100%	100%	100%	99%
0.4	99%	99%	98%	99%
0.5	96%	100%	97%	99%
0.6	97%	96%	98%	99%
0.7	98%	91%	95%	88%
0.8	98%	56%	30%	28%
0.9	97%	0%	0%	0%
1.0	0%	0%	0%	0%

APPENDIX A.3 Table 1, Sunglasses - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	100%	87%	89%	83%
1	90%	90%	0%	88%
2	95%	100%	0%	90%
3	96%	100%	0%	95%
4	99%	98%	0%	95%
5	98%	99%	0%	96%
6	93%	97%	0%	97%
7	96%	98%	0%	98%
8	95%	95%	0%	98%
9	95%	96%	0%	99%
10	91%	55%	0%	100%

APPENDIX A.3 Table 2, Sunglasses - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	98%	85%	88%	82%
10	100%	84%	0%	85%
20	100%	90%	0%	83%
30	98%	91%	0%	84%
40	99%	95%	0%	86%
50	98%	97%	0%	90%
60	100%	98%	0%	80%
70	100%	80%	0%	30%
80	100%	12%	0%	10%
90	99%	0%	0%	0%
100	88%	0%	0%	0%

APPENDIX A.3 Table 3, Sunglasses - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	98%	96%	87%	82%
0.1	100%	96%	90%	92%
0.2	98%	98%	88%	98%
0.3	97%	85%	0%	95%
0.4	98%	97%	0%	97%
0.5	99%	98%	0%	96%
0.6	97%	99%	0%	98%
0.7	98%	99%	0%	99%
0.8	99%	98%	0%	97%
0.9	100%	97%	0%	95%
1.0	0%	0%	0%	0%

APPENDIX A.4 Table 1, Electric switch - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	100%	99%	93%	99%
1	100%	100%	95%	99%
2	100%	100%	97%	100%
3	98%	100%	97%	100%
4	97%	98%	94%	98%
5	95%	88%	95%	96%
6	85%	82%	88%	97%
7	81%	78%	70%	98%
8	63%	68%	60%	98%
9	65%	69%	28%	99%
10	45%	55%	11%	100%

APPENDIX A.4 Table 2, Electric switch - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	100%	99%	94%	98%
10	99%	100%	90%	97%
20	100%	100%	84%	96%
30	98%	100%	84%	93%
40	99%	100%	73%	86%
50	98%	92%	40%	82%
60	99%	85%	18%	75%
70	93%	74%	0%	30%
80	92%	36%	0%	0%
90	83%	14%	0%	0%
100	31%	0%	0%	0%

APPENDIX A.4 Table 3, Electric switch - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	99%	100%	97%	99%
0.1	99%	99%	95%	100%
0.2	100%	100%	93%	100%
0.3	99%	86%	87%	98%
0.4	100%	82%	67%	97%
0.5	100%	83%	48%	98%
0.6	99%	74%	38%	98%
0.7	99%	62%	38%	99%
0.8	99%	38%	0%	97%
0.9	100%	26%	0%	95%
1.0	0%	0%	0%	0%

APPENDIX A.5 Table 1, Envelope - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	99%	90%	83%	99%
1	99%	92%	82%	99%
2	99%	95%	82%	99%
3	76%	85%	70%	90%
4	67%	74%	71%	80%
5	40%	60%	46%	69%
6	42%	40%	33%	43%
7	23%	19%	13%	17%
8	16%	0%	11%	16%
9	10%	0%	12%	12%
10	0%	0%	11%	0%

APPENDIX A.5 Table 2, Envelope - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	100%	93%	80%	98%
10	98%	94%	81%	96%
20	98%	95%	83%	95%
30	95%	97%	76%	88%
40	81%	85%	58%	71%
50	75%	45%	40%	58%
60	31%	24%	18%	20%
70	0%	0%	0%	0%
80	0%	0%	0%	0%
90	0%	0%	0%	0%
100	0%	0%	0%	0%

APPENDIX A.5 Table 3, Envelope - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	99%	98%	82%	99%
0.1	99%	99%	85%	99%
0.2	100%	100%	83%	95%
0.3	100%	99%	85%	96%
0.4	100%	99%	80%	91%
0.5	99%	98%	77%	92%
0.6	99%	96%	71%	83%
0.7	86%	65%	64%	84%
0.8	47%	19%	30%	30%
0.9	35%	10%	0%	12%
1.0	0%	0%	0%	0%

APPENDIX A.6 Table 1, Toilet seat - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	99%	100%	92%	98%
1	98%	100%	91%	94%
2	99%	100%	82%	92%
3	98%	100%	71%	90%
4	99%	98%	71%	91%
5	97%	99%	24%	91%
6	94%	97%	0%	93%
7	94%	98%	0%	91%
8	95%	99%	0%	83%
9	95%	100%	0%	58%
10	92%	97%	0%	10%

APPENDIX A.6 Table 2, Toilet seat - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	98%	99%	93%	97%
10	98%	100%	86%	96%
20	98%	99%	84%	95%
30	95%	99%	76%	88%
40	95%	95%	66%	84%
50	96%	88%	37%	81%
60	85%	70%	0%	65%
70	54%	51%	0%	40%
80	46%	29%	0%	10%
90	13%	12%	0%	0%
100	0%	0%	0%	0%

APPENDIX A.6 Table 3, Toilet seat - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	99%	100%	89%	99%
0.1	99%	100%	91%	99%
0.2	98%	100%	81%	98%
0.3	99%	100%	0%	96%
0.4	99%	99%	0%	95%
0.5	99%	99%	0%	94%
0.6	99%	100%	0%	94%
0.7	95%	98%	0%	88%
0.8	96%	93%	0%	86%
0.9	84%	92%	0%	74%
1.0	0%	0%	0%	0%

APPENDIX A.7 Table 1, Plastic bag - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	97%	89%	86%	94%
1	96%	83%	80%	93%
2	98%	84%	76%	92%
3	98%	79%	69%	93%
4	96%	65%	52%	94%
5	95%	68%	43%	93%
6	88%	59%	0%	87%
7	81%	51%	0%	77%
8	60%	46%	0%	47%
9	43%	41%	0%	0%
10	12%	37%	0%	0%

APPENDIX A.7 Table 2, Plastic bag - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	98%	87%	88%	95%
10	95%	84%	79%	93%
20	95%	78%	61%	84%
30	91%	71%	10%	82%
40	93%	62%	0%	76%
50	96%	50%	0%	46%
60	85%	48%	0%	34%
70	79%	40%	0%	20%
80	46%	32%	0%	0%
90	13%	14%	0%	0%
100	0%	0%	0%	0%

APPENDIX A.7 Table 3, Plastic bag - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	97%	83%	85%	96%
0.1	98%	84%	82%	97%
0.2	97%	82%	83%	98%
0.3	98%	82%	84%	98%
0.4	96%	78%	65%	97%
0.5	95%	79%	48%	98%
0.6	97%	74%	38%	94%
0.7	95%	66%	23%	88%
0.8	96%	56%	11%	77%
0.9	84%	26%	0%	74%
1.0	0%	0%	0%	0%

APPENDIX A.8 Table 1, Soup bowl - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	99%	98%	89%	93%
1	99%	99%	16%	93%
2	98%	99%	0%	94%
3	99%	100%	0%	95%
4	98%	97%	0%	95%
5	98%	100%	0%	89%
6	98%	100%	0%	87%
7	99%	100%	0%	77%
8	95%	95%	0%	54%
9	98%	96%	0%	43%
10	91%	58%	0%	40%

APPENDIX A.8 Table 2, Soup bowl - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	98%	98%	88%	95%
10	98%	98%	81%	96%
20	98%	100%	54%	92%
30	100%	97%	39%	87%
40	99%	95%	0%	83%
50	99%	83%	0%	80%
60	99%	70%	0%	35%
70	100%	58%	0%	0%
80	100%	32%	0%	0%
90	99%	14%	0%	0%
100	88%	0%	0%	0%

APPENDIX A.8 Table 3, Soup bowl - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	98%	98%	87%	92%
0.1	98%	96%	90%	96%
0.2	98%	98%	88%	98%
0.3	97%	99%	85%	98%
0.4	97%	99%	60%	97%
0.5	98%	98%	32%	96%
0.6	97%	93%	0%	95%
0.7	96%	69%	0%	96%
0.8	96%	19%	0%	91%
0.9	97%	0%	0%	89%
1.0	0%	0%	0%	0%

APPENDIX A.9 Table 1, Band-aid - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	97%	100%	92%	93%
1	95%	100%	91%	94%
2	93%	100%	87%	92%
3	94%	100%	71%	90%
4	96%	99%	67%	83%
5	97%	100%	37%	84%
6	98%	100%	24%	93%
7	99%	100%	13%	91%
8	95%	99%	0%	83%
9	98%	100%	0%	58%
10	92%	97%	0%	40%

APPENDIX A.9 Table 2, Band-aid - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	95%	99%	93%	94%
10	96%	100%	86%	96%
20	97%	100%	84%	92%
30	93%	100%	74%	87%
40	95%	100%	66%	84%
50	90%	88%	24%	81%
60	79%	85%	0%	75%
70	54%	74%	0%	40%
80	27%	53%	0%	0%
90	0%	42%	0%	0%
100	0%	19%	0%	0%

APPENDIX A.9 Table 3, Band-aid - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	96%	100%	89%	92%
0.1	98%	100%	91%	96%
0.2	96%	100%	81%	94%
0.3	93%	100%	77%	92%
0.4	96%	99%	60%	93%
0.5	94%	95%	32%	94%
0.6	86%	93%	12%	84%
0.7	83%	92%	0%	77%
0.8	96%	76%	0%	86%
0.9	97%	0%	0%	0%
1.0	0%	0%	0%	0%

APPENDIX A.10 Table 1, Dishwasher - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	90%	95%	86%	94%
1	90%	96%	84%	90%
2	93%	100%	87%	92%
3	71%	99%	70%	90%
4	67%	96%	71%	67%
5	61%	87%	38%	56%
6	61%	79%	31%	36%
7	58%	72%	13%	18%
8	54%	63%	0%	14%
9	53%	61%	0%	0%
10	47%	58%	0%	0%

APPENDIX A.10 Table 2, Dishwasher - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	98%	96%	88%	94%
10	95%	94%	84%	96%
20	95%	90%	83%	80%
30	95%	91%	74%	63%
40	97%	85%	58%	39%
50	98%	70%	24%	37%
60	89%	60%	0%	34%
70	90%	40%	0%	0%
80	79%	12%	0%	0%
90	50%	0%	0%	0%
100	46%	0%	0%	0%

APPENDIX A.10 Table 3, Dishwasher - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	97%	96%	84%	92%
0.1	98%	100%	82%	97%
0.2	97%	100%	83%	95%
0.3	98%	100%	76%	95%
0.4	98%	99%	43%	93%
0.5	100%	98%	16%	93%
0.6	98%	95%	12%	83%
0.7	95%	92%	0%	84%
0.8	97%	68%	0%	77%
0.9	63%	10%	0%	12%
1.0	0%	0%	0%	0%

APPENDIX A.11 Table 1, Watch - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	90%	79%	60%	60%
1	74%	70%	16%	53%
2	75%	46%	0%	51%
3	73%	35%	0%	34%
4	63%	21%	0%	25%
5	30%	14%	0%	24%
6	12%	11%	0%	22%
7	0%	10%	0%	18%
8	0%	0%	0%	14%
9	0%	0%	0%	11%
10	0%	0%	0%	0%

APPENDIX A.11 Table 2, Watch - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	95%	86%	68%	60%
10	83%	80%	11%	48%
20	78%	68%	0%	26%
30	64%	42%	0%	21%
40	45%	36%	0%	0%
50	14%	15%	0%	0%
60	0%	5%	0%	0%
70	0%	0%	0%	0%
80	0%	0%	0%	0%
90	0%	0%	0%	0%
100	0%	0%	0%	0%

APPENDIX A.11 Table 3, Watch - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	93%	74%	68%	56%
0.1	86%	86%	14%	62%
0.2	83%	85%	0%	53%
0.3	82%	85%	0%	54%
0.4	80%	82%	0%	55%
0.5	81%	83%	0%	56%
0.6	81%	76%	0%	55%
0.7	79%	62%	0%	54%
0.8	46%	33%	0%	36%
0.9	13%	0%	0%	0%
1.0	0%	0%	0%	0%

APPENDIX A.12 Table 1, Pill bottle - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	96%	90%	82%	92%
1	74%	90%	0%	91%
2	71%	95%	0%	90%
3	76%	83%	0%	90%
4	64%	80%	0%	83%
5	30%	82%	0%	84%
6	42%	76%	0%	42%
7	23%	65%	0%	17%
8	16%	48%	0%	16%
9	10%	40%	0%	11%
10	0%	37%	0%	0%

APPENDIX A.12 Table 2, Pill bottle - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	95%	93%	82%	93%
10	96%	93%	56%	93%
20	97%	95%	43%	90%
30	91%	77%	10%	84%
40	93%	68%	0%	76%
50	90%	67%	0%	58%
60	79%	54%	0%	54%
70	79%	26%	0%	20%
80	27%	0%	0%	0%
90	0%	0%	0%	0%
100	0%	0%	0%	0%

APPENDIX A.12 Table 3, Pill bottle - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	98%	94%	82%	96%
0.1	98%	96%	83%	97%
0.2	98%	89%	83%	94%
0.3	98%	86%	84%	95%
0.4	96%	97%	80%	92%
0.5	99%	95%	77%	92%
0.6	95%	90%	71%	84%
0.7	86%	66%	64%	77%
0.8	83%	33%	46%	36%
0.9	81%	0%	0%	0%
1.0	0%	0%	0%	0%

APPENDIX A.13 Table 1, Toilet paper - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	96%	95%	93%	92%
1	95%	96%	96%	91%
2	94%	94%	97%	92%
3	95%	85%	97%	90%
4	92%	80%	94%	91%
5	93%	77%	95%	89%
6	89%	76%	88%	86%
7	87%	73%	70%	77%
8	88%	68%	60%	54%
9	83%	69%	36%	43%
10	84%	58%	10%	11%

APPENDIX A.13 Table 2, Toilet paper - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	96%	96%	94%	93%
10	98%	93%	95%	96%
20	98%	82%	96%	96%
30	98%	77%	95%	93%
40	99%	68%	80%	83%
50	98%	70%	81%	82%
60	89%	60%	71%	65%
70	90%	51%	50%	52%
80	79%	36%	30%	33%
90	64%	12%	10%	13%
100	30%	0%	0%	0%

APPENDIX A.13 Table 3, Toilet paper - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	97%	94%	97%	92%
0.1	95%	96%	98%	97%
0.2	98%	89%	99%	91%
0.3	98%	86%	100%	92%
0.4	97%	87%	98%	92%
0.5	96%	87%	97%	93%
0.6	95%	90%	98%	95%
0.7	96%	91%	95%	96%
0.8	98%	93%	46%	91%
0.9	97%	92%	0%	89%
1.0	0%	0%	0%	0%

APPENDIX A.14 Table 1, Remote controller - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	96%	98%	55%	98%
1	99%	99%	54%	93%
2	99%	99%	50%	94%
3	98%	99%	55%	95%
4	98%	97%	41%	82%
5	88%	87%	38%	71%
6	85%	83%	13%	36%
7	70%	72%	10%	0%
8	63%	61%	0%	0%
9	65%	40%	0%	0%
10	45%	24%	0%	0%

APPENDIX A.14 Table 2, Remote controller - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	98%	98%	60%	97%
10	94%	98%	56%	97%
20	95%	99%	43%	90%
30	95%	100%	39%	63%
40	97%	95%	0%	24%
50	98%	92%	0%	11%
60	99%	54%	0%	0%
70	99%	14%	0%	0%
80	92%	0%	0%	0%
90	83%	0%	0%	0%
100	46%	0%	0%	0%

APPENDIX A.14 Table 3, Remote controller - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	98%	96%	50%	99%
0.1	97%	99%	51%	99%
0.2	98%	99%	49%	99%
0.3	96%	100%	42%	98%
0.4	96%	99%	31%	95%
0.5	98%	100%	16%	93%
0.6	91%	100%	10%	75%
0.7	95%	98%	0%	71%
0.8	83%	66%	0%	28%
0.9	63%	0%	0%	0%
1.0	0%	0%	0%	0%

APPENDIX A.15 Table 1, Soap dispenser - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	97%	89%	83%	86%
1	96%	92%	82%	88%
2	94%	94%	76%	89%
3	95%	85%	58%	81%
4	96%	78%	52%	82%
5	93%	77%	34%	71%
6	89%	59%	9%	73%
7	96%	51%	10%	49%
8	89%	43%	0%	47%
9	83%	20%	0%	14%
10	84%	0%	0%	0%

APPENDIX A.15 Table 2, Soap dispenser - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	96%	87%	80%	90%
10	96%	78%	79%	91%
20	95%	68%	61%	84%
30	93%	70%	46%	82%
40	81%	64%	0%	71%
50	75%	45%	0%	46%
60	50%	48%	0%	20%
70	27%	14%	0%	0%
80	0%	0%	0%	0%
90	0%	0%	0%	0%
100	0%	0%	0%	0%

APPENDIX A.15 Table 3, Soap dispenser - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	93%	96%	85%	88%
0.1	86%	100%	85%	90%
0.2	82%	99%	80%	91%
0.3	82%	100%	77%	92%
0.4	80%	99%	65%	91%
0.5	81%	98%	65%	93%
0.6	81%	99%	50%	75%
0.7	72%	99%	23%	71%
0.8	55%	98%	11%	30%
0.9	54%	97%	0%	12%
1.0	0%	0%	0%	0%

APPENDIX A.16 Table 1, Banana - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	99%	99%	0%	91%
1	98%	100%	0%	90%
2	96%	100%	0%	89%
3	93%	99%	0%	79%
4	67%	98%	0%	67%
5	50%	94%	0%	56%
6	46%	55%	0%	42%
7	40%	50%	0%	23%
8	37%	43%	0%	24%
9	23%	15%	0%	14%
10	0%	10%	0%	0%

APPENDIX A.16 Table 2, Banana - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	98%	99%	0%	90%
10	96%	100%	0%	91%
20	95%	99%	0%	80%
30	76%	62%	0%	32%
40	29%	0%	0%	0%
50	0%	0%	0%	0%
60	0%	0%	0%	0%
70	0%	0%	0%	0%
80	0%	0%	0%	0%
90	0%	0%	0%	0%
100	0%	0%	0%	0%

APPENDIX A.16 Table 3, Banana - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	99%	100%	0%	88%
0.1	98%	100%	0%	90%
0.2	98%	100%	0%	94%
0.3	99%	100%	0%	95%
0.4	99%	99%	0%	89%
0.5	100%	99%	0%	82%
0.6	98%	74%	0%	45%
0.7	72%	0%	0%	0%
0.8	24%	0%	0%	0%
0.9	0%	0%	0%	0%
1.0	0%	0%	0%	0%

APPENDIX A.17 Table 1, Television - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	97%	89%	60%	83%
1	95%	83%	54%	83%
2	96%	46%	50%	83%
3	93%	35%	55%	79%
4	92%	21%	41%	80%
5	88%	14%	43%	69%
6	88%	11%	24%	61%
7	70%	10%	13%	47%
8	60%	0%	11%	17%
9	43%	0%	12%	12%
10	12%	0%	0%	11%

APPENDIX A.17 Table 2, Television - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	98%	86%	68%	86%
10	94%	84%	0%	85%
20	95%	82%	0%	83%
30	64%	42%	0%	42%
40	29%	36%	0%	0%
50	14%	15%	0%	0%
60	0%	5%	0%	0%
70	0%	0%	0%	0%
80	0%	0%	0%	0%
90	0%	0%	0%	0%
100	0%	0%	0%	0%

APPENDIX A.17 Table 3, Television - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	97%	83%	68%	85%
0.1	97%	82%	51%	92%
0.2	98%	82%	49%	94%
0.3	93%	82%	42%	92%
0.4	96%	78%	31%	89%
0.5	94%	79%	20%	82%
0.6	91%	76%	10%	45%
0.7	79%	65%	0%	0%
0.8	46%	66%	0%	0%
0.9	0%	26%	0%	0%
1.0	0%	0%	0%	0%

APPENDIX A.18 Table 1, Washing machine - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	96%	87%	55%	91%
1	95%	84%	80%	93%
2	95%	84%	69%	92%
3	94%	85%	69%	93%
4	67%	74%	67%	78%
5	50%	68%	46%	68%
6	46%	40%	33%	43%
7	40%	19%	32%	0%
8	37%	0%	35%	0%
9	23%	0%	36%	0%
10	0%	0%	10%	0%

APPENDIX A.18 Table 2, Washing machine - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	95%	85%	60%	82%
10	82%	84%	11%	85%
20	78%	80%	0%	66%
30	76%	62%	0%	32%
40	45%	0%	0%	0%
50	0%	0%	0%	0%
60	0%	0%	0%	0%
70	0%	0%	0%	0%
80	0%	0%	0%	0%
90	0%	0%	0%	0%
100	0%	0%	0%	0%

APPENDIX A.18 Table 3, Washing machine - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	96%	86%	50%	82%
0.1	95%	84%	14%	83%
0.2	96%	85%	0%	84%
0.3	96%	85%	0%	64%
0.4	96%	83%	0%	47%
0.5	95%	83%	0%	46%
0.6	86%	74%	0%	20%
0.7	83%	76%	0%	15%
0.8	24%	76%	0%	14%
0.9	13%	0%	0%	0%
1.0	0%	0%	0%	0%

APPENDIX A.19 Table 1, Coffee mug - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	85%	89%	82%	86%
1	83%	84%	84%	83%
2	71%	79%	69%	83%
3	71%	83%	58%	81%
4	64%	78%	50%	78%
5	61%	82%	34%	68%
6	61%	79%	31%	61%
7	58%	78%	32%	47%
8	54%	63%	35%	25%
9	53%	61%	28%	14%
10	47%	58%	0%	10%

APPENDIX A.19 Table 2, Coffee mug - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	82%	80%	82%	86%
10	82%	78%	84%	85%
20	74%	80%	54%	66%
30	71%	70%	46%	42%
40	59%	64%	0%	39%
50	51%	67%	0%	37%
60	50%	70%	0%	35%
70	27%	58%	0%	0%
80	13%	29%	0%	0%
90	0%	14%	0%	0%
100	0%	0%	0%	0%

APPENDIX A.19 Table 3, Coffee mug - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	84%	86%	84%	85%
0.1	83%	82%	83%	83%
0.2	82%	85%	80%	84%
0.3	76%	86%	76%	64%
0.4	72%	83%	67%	47%
0.5	73%	83%	65%	46%
0.6	66%	84%	50%	20%
0.7	65%	76%	38%	15%
0.8	55%	68%	0%	14%
0.9	54%	26%	0%	12%
1.0	0%	0%	0%	0%

APPENDIX A.20 Table 1, Pillow - Blur distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	85%	79%	0%	60%
1	83%	70%	0%	53%
2	75%	79%	0%	51%
3	73%	79%	0%	34%
4	63%	65%	0%	25%
5	40%	60%	0%	24%
6	12%	55%	0%	22%
7	0%	50%	0%	23%
8	0%	46%	0%	24%
9	0%	15%	0%	0%
10	0%	10%	0%	0%

APPENDIX A.20 Table 2, Pillow - Noise distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0	82%	80%	0%	60%
10	83%	80%	0%	48%
20	74%	78%	0%	26%
30	71%	71%	0%	21%
40	59%	62%	0%	24%
50	51%	50%	0%	11%
60	31%	24%	0%	0%
70	0%	0%	0%	0%
80	13%	0%	0%	0%
90	0%	0%	0%	0%
100	0%	0%	0%	0%

APPENDIX A.20 Table 3, Pillow - Contrast distortion impact

Distortion	Inception v3	ResNet50	SqueezeNet	MobileNet
0.0	84%	74%	0%	56%
0.1	83%	86%	0%	62%
0.2	83%	85%	0%	53%
0.3	76%	85%	0%	54%
0.4	72%	87%	0%	55%
0.5	73%	87%	0%	56%
0.6	66%	84%	0%	55%
0.7	65%	0%	0%	54%
0.8	47%	0%	0%	36%
0.9	35%	0%	0%	0%
1.0	0%	0%	0%	0%

