Turkish Journal of Computer and Mathematics Education *Volume 15, Issue 1, 169-182*
Research Article

Depth reduction of RGB image data and reduction of point noise based on metric learning method.

Riyadh Alsaeedi [a]

[a] Wasit University, Iraq

Abstract: In this paper, a method of data depth reduction based on metric learning method in reducing point noise in different images is proposed. In order to be more accurate in reviving depth from data, noise variance is also calculated for each separate scale. In this way, our method becomes more sensitive to noise detection. The quantitative and qualitative results obtained from the implementation and calculation of the PSNR parameter of this method show that the proposed method of this paper has given a good answer compared to previous methods for noise elimination and has performed better in maintaining sharp corners and sharp features.

Keywords: Depth Reduction, RGB Image Data, Noise Reduction, Metric Learning, Image Processing Techniques

Introduction

Applications for pattern recognition and machine learning, such as face recognition, scene recognition, etc., are affected by the quality of the input data. Therefore, it is important to quantitatively evaluate the quality of the images, which reflects the image's ability to be used as a good example. Considering the growth and widespread presence of digital images everywhere, it is essential to have efficient and reliable methods for evaluating image quality.

The quality of an image sample is subject to a variety of distortions during recording, storage, or transfer, which may reduce image quality. The purpose of image quality assessment methods is to automatically check the quality of images in line with human opinion. In measuring image quality, a person can easily give a mental score to the observed image. However, embedding such a mechanism in computer vision systems or image processing is a difficult task. Therefore, the method of image quality assessment that automatically produces an objective measurement consistent with human mental outcomes can be very desirable. In general, the purpose of image quality evaluation algorithms is to assess the perceptual quality of an image using objective indicators that are highly consistent with the human Subjective index.

Image quality evaluation algorithms can be divided into three areas: With a Full-Reference with a Reduced Reference and No-Reference. The first two domains require complete or partial reference image information, while referenceless algorithms do not require or access any information. As a result, a reference (or blind) image quality evaluation algorithm is preferred in situations where access to any reference information is impossible.

Non-reference algorithms are divided into two main subgroups according to previous knowledge of distortion type. Distortion Specific (DS), No-Distortion –Specific (NDS) The DS NR-IQA assumes that the type of distortion that affects the image is clear, which is measured separately from other factors. Compared to DS NR-IQA, previous knowledge of the type of distortion by NDS NR-IQA algorithms has not been considered. Instead, the quality score is based on the assumption that the image is evaluated by the type of distortion similar to the data in the training database.

Many NDS NR-IQA algorithms follow one of two approaches: (1) the method of natural scene statistics (NSS) and (2) the approach based on learning or teaching.

Algorithms that use natural image statistics examine the statistical state of the naturalness of distorted images. Gabarda and Cristóbal [1] used Anisotropy as a measure of image quality. Heterogeneity means having different characteristics in different directions. Generalized Renyi entropy [2] and Pseudo-Wigner distribution normalized PWD-like distribution (PWD) have been used to calculate the directional entropy of an image.

Lu et al. [3] proposed a contour-based natural scene statistical (NSS) based on contour transform for image quality assessment. Cantorlet coefficient statistics are described by the Joint distribution function. In this method, the image is decomposed using a controlling converter under multi-scale and multi-directional bands. Moorthy and Bovik [4] provided a two-step framework for assessing image quality. This algorithm is called the BIQI. This method first estimates the presence of a set of distortions and then uses the weighted probability sum to obtain the quality score.

Wave conversion is applied on three scales and three directions to obtain sub-band coefficients. These coefficients are parameterized using the generalized Gaussian distribution. Feature vectors obtained from sub-band coefficients are given to a support vector machine (SVM) to classify different types of distortion.

Moorthy and Bovik [5] proposed a Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE). DIIVINE is based on the hypothesis that the statistical characteristics of images change in the presence of distortion. In this algorithm, the distorted image is decomposed into waves to obtain transit responses. Dividing the wave of a guided pyramid [58] (guided filters are convention-selectable filters) on two scales (1, 2) and six directions (0◦, 30◦, 60◦, 90◦, 120, 120) ◦, 150◦) is used. Subtraction coefficients are then used to extract statistical characteristics.

Saad et al. [6] proposed a blind memory reminder algorithm using DCT statistics (BLIINDS)[1]. This algorithm uses discrete cosine transform (DCT) to extract the structural and contrast properties of an image. Contrasts are obtained from the average value of patch DCT coefficients (blocks) with dimensions of 17. 17. Kurtosis marks (the probability of peak distribution distribution) of squares of the same size are used to obtain structural features based on DCT.

Mittal et al. [7] proposed mean subtracted contrast normalized or BRISQUE as a metric for image quality in the field of location. This algorithm uses locally normalized luminance brightness [26], mean subtracted contrast normalized (MSCN) of the image (I ^ '). The algorithm applies generalized Gaussian distribution to obtain distortion image statistics that tend to show a change in the distribution of coefficients in the presence of distortion.
Compared to statistical methods based on natural image statistics, in which appropriate characteristics are determined using the statistical status of natural images, in methods based on learning, features are discovered through the machine learning process. Li et al. [8] presented an approach based on general regression of the neural network to assess image quality. The fusion image phase (Congruency) [50], the entropy image phase entropy, and the distorted image gradient are used as attributes. These characteristics and the average differential score (DMOS), ie the Subjective rankings of the image (based on human opinion), are given to obtain the connection to the neural network.
Ye and Doermann [9] provided a visual image metric based on visual Codebook. Codebook means a set of specific features that determine the quality of an image. The Gabor filter is used in five frequencies and four directions in size of 8. 8image blocks. The mean and variance of the output filter are used as a feature vector. The k-means cluster is used to create 200 clusters corresponding to the educational image.
Kang et al. [10] suggested using a shallow convection neural network as a pioneer. The proposed structure consists of 4 layers (a convolutional layer and three complete connection layers) and an input with a size of 32. 32.

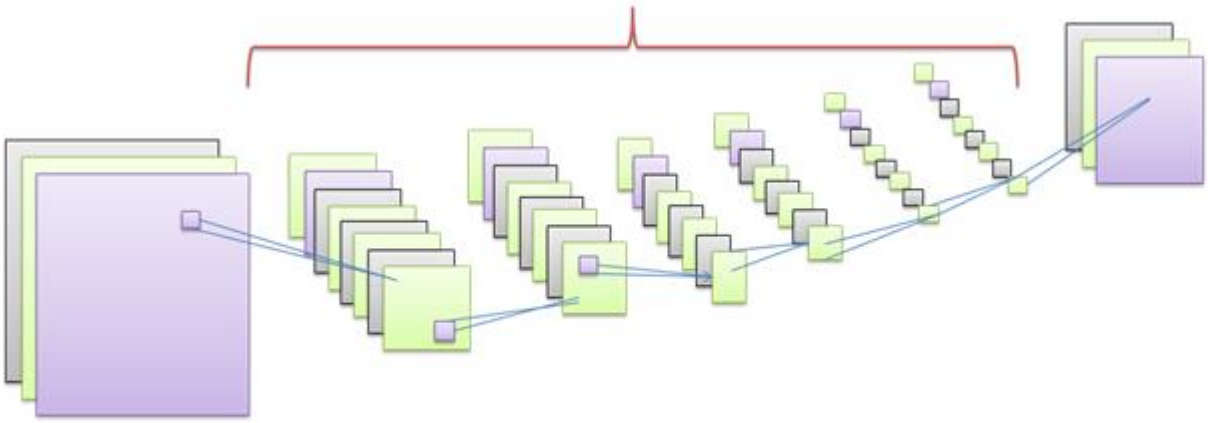
Bianco et al. [11] used Pre-trained convolutional neural networks as image descriptors to extract the feature and fine-tuned to assess image quality. They estimate the image quality by averaging the predicted scores of different parts of an image. To predict the quality of each piece, they used the image of the support vector regression.
Zhang et al. [12] suggested learning the stereoscopic structure based on convolutional neural networks. Image blocks are created. Outstanding images are given to the network as input, which allows the network to learn local structures that are sensitive to human perception and act as a measure of perceptual quality. They designed two types of networks, a single input network and a three-input network with three different views of the image. Their proposed network with a few changes can be used to evaluate the quality of images without reference or with reference.
Boss et al. [13] proposed a network with ten convolutional layers, five pooling layers, and two complete connection layers for regression. Their proposed network with a few changes can be used to evaluate the quality of images without reference or with reference. It calculates the network result for different parts of the image and calculates the overall quality by averaging them.

Methods based on statistical images of natural images are highly dependent on the handmade feature (which is based on predefined algorithms and empirical knowledge). It is also expensive to transfer images to a new domain. Learning-based and Codebook methods are more successful if they use large codecs in the feature extraction phase. At the same time, the independence of the extraction feature and the training of the regression model in this type of algorithms is somewhat challenging. Deep learning-based methods assume that the distortion is homogeneously distributed throughout the image. The input of these methods is small squares (blocks) of the input image, which are often in the size of 32 * 32 or 7 * 7; in some cases, the distortions are smaller. In this case, the quality score of the original image is calculated based on small squares (blocks). Although this method increases the size of the training dataset, it is

[1] Blind Image Integrity Notator using DCT Statistics

perceptually wrong; because based on the structural and content information in the squares, each of the image squares (local information) can have different quality scores, which is completely different from the final image quality. Also, the convolutional networks designed to assess image quality use a maximum of three convolution layers to extract the feature, which prevents them from learning the appropriate distinguishing features from the input image. Due to the fact that these algorithms are highly dependent on the training data, when the number and type of distortions increase, the efficiency of the algorithms is affected compared to the number of low distortions.
As shown in Figure (1), CNN consists of a stack of convolution layers and then completely interconnected layers. This architecture consists of 5 layers of convolution, the first and second layers being separated by the max-pooling layer. The five convolution layers are followed by the intermediate merging layer that follows the softmax function.

Input **Features structure** **Output**



12* 12 convolution 5 * 5 convolution 3 * 3 convolution 3 * 3 convolution 3 * 3 convolution 3 * 3 convolution

Figure 1: Recommended Convolutional Neural Network

Usually, the convolution layer includes different operators such as convolution, nonlinear conversion, and integration. The result of the merger is a new image (called a feature map) in which each element of the calculations is obtained at a point in the local area that is connected to the input feature map and a set of weights (filters or cores). This operator is followed by a nonlinear function (such as ReLU, tanh, etc.) and then by a merger function. Integration, through a predefined function, samples local areas among the spatial dimensions of feature maps.

The completely cohesive layer takes all the neurons from the previous layer (whether the convolution layer or any other completely cohesive layer) and connects them to a neuron in its layer. Usually by dropout regularization method used to reduce overfitting in layers, and accidentally removes several neurons that do not play an effective role in the network.

The completely correlated layer output is used to generate the potential output of each layer (layer layer). The most common conversion function for multi-class classification is the softmax function. This function is the vector that shows the probability of classification for each class in the range (1.0) of the distribution.

Activate and merge function

The image convolution layer with W × H path and N channel with x (i, j) center and 2D filter core, wf w hf as input and output feature map (H-hf + 1) × (W-wf + 1) Produces the channel with k. Each channel of the output image is a filter. The output of the convolution process is affected by the sf parameter. Sf The distance required for the convolution process in the input image or map is a feature. If 1 <sf is the size of an output map from the convolution process, it will be reduced to (W-wf) / sf + 1) × ((H - hf) / sf + 1).The convolution process is defined as a relation (3-1):

$$(1\text{-}3) \qquad \sum_{n=1}^{N} \left\{ \sum_{p=0}^{w_f-1} \sum_{q=0}^{h_f-1} x_n \left(i.s_f + p, j.s_f + q\right).h_k(p,q) \right\} + b_k$$

xn (i, j), xk (ii, jj), hk (p, q) and bk The amount of pixels (i, j) in the nm channel is an input image or feature map.

The convolution and conversion operator is called the activation function.We assume that the input xk (ii, jj) is a function of the activation of the neural network, which is the output of the convolution process. W is the weight vector and b is the bias vector. The activation function is expressed as a relation (3-2):

$$(2\text{-}3) \quad Z(x_k(ii,jj) = f\left(\sum_{k=1}^{k} x_k(ii,jj).w_k + b_k\right) \quad \Leftrightarrow \quad Z = f(X.W + b)$$

There are many alternative functions such as tanh, sigmoid, hyperbolic, and modified functions for f (0). Currently, the modified function is more commonly used for neurons because it prevents saturation during the metric learning process and does not cause gradient problems. This research uses the ReLU[1] function:

$$(3\text{-}3) \qquad A(x_k(ii,jj) = \max\left(0, Z\left(x_k(ii,jj)\right)\right)$$

The merger operator performs spatial sampling by considering the maximum or average value of the wp × hp merger window. This operator ensures that similar results can be obtained even when the image features are rotating. If we assume that A (xk (ii, jj)) is the output of the activation operator and uses max-pooling with sp to express the output of xk (iip, jjp) as follows:

$$(4\text{-}3) \qquad x_k\left(ii_p, jj_p\right) = max_{0 \le ii_p \le h_p - 1, 0 \le jj_p \le w_p - 1} A(x(ii,jj))$$

In the max-pooling process, the input of the kM channel with the size ((W-wf) / sf + 1) × ((W-wf) / sf + 1) to ((W-wf) /sf.sp+1) × (W-wf) /sf.sp+1) decreases. The main idea of the max-pooling operator is shown in Figure (2).

[1] Rectified Linear Unit

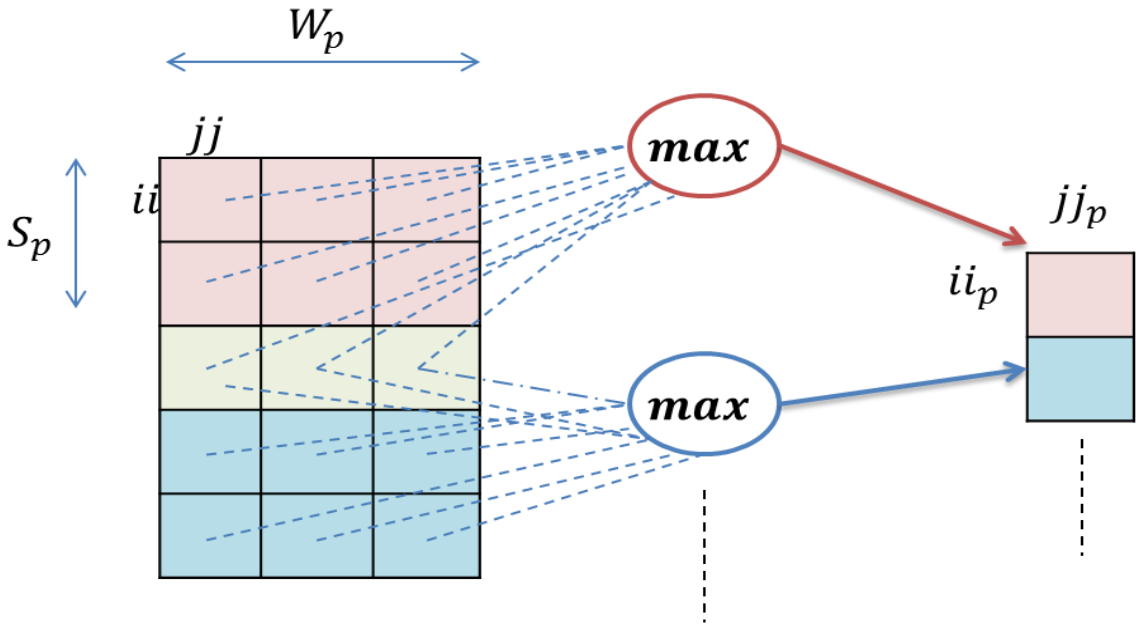


Figure 2: The main idea of the max-pooling operator

Classification

After the convolutional and fully cohesive layers, the stratified layer is used to predict the probabilities of the stratum, which is represented as the multi-channel m path of the input image path and the ground-truth image path. The most common conversion function for multimax proof is the softmax function. Assume that Wm × Hm × k is the modified output of CNN, that K is the number of channels in the output image path path. X = [X1,…, XK] T shows the output of the fully correlated layer, and the softmax function for converting x to the probability vector m = [m1,…, mk] T is as follows:

$$(5\text{-}3) \quad m_{w,b} = \frac{\exp\,(x.w_c)}{\sum_k^K \exp\,(\mathrm{x}.w_k)}$$

Figure (3-4) shows the proposed CNN architecture. This architecture consists of 5 layers of convolution, the first and second layers being separated by the max-pooling layer. The five convolution layers are followed by the intermediate merging layer that follows the softmax function.

In the same way as in [5], the input image is divided into 64 × 64 paths, three channels (red, green, and blue); The output is a 768-dimensional vector that is transformed into three 16 × 16 channels (road, depth of RGB-D data, and background). Decision-making in choosing the input path is wider than the output path, which can be used to provide more information to predict the final probability map, thus identifying larger classes (such as the depth of RGB-D data).) To be simpler. If cohesion is used effectively in CNNs, the decision to choose a multi-class forecasting strategy will require more precision than single-class forecasting. Each input path is normalized by subtracting the mean value and dividing it by the standard deviation.

The last layer of convolution is followed by the integration of the global average pooling (GAP). When we expect similar results along the way, GAP simply averages feature maps. However, the completely traditional bonding layer is the result of mapping the feature of the last layer of the convolution layer to a layer. The fully bonded layer suffers from post-fit parameters, which require a lot of shrinkage to reduce it. So that half or more of the activities of the completely cohesive layers during the training are reduced to zero and require optimization in setting the parameters. However, GAP does not require any optimization.

As shown in Figure (2), the input path is 64 × 64 @ 3, which includes three channels with 64 × 64 dimensions. The first convolution layer is 14 × 14 @ 128, which consists of 128 filters-channels and each has dimensions of 14 × 14. It results in 13 × 13 @ 128. This process is followed by the 13 × 13 @ 128 concentration along with the 5 × 5, 3 × 3, 3 × 3, and 3 × 3 filter cores with filter units 256, 512, 32, and 768, and then outputs, respectively. It produces with dimensions of 9 × 9 @ 256, 7 × 7 @ 512, 5 × 5 3 32 and 3 × 3. 768. All convolution layers, except the first layer, which has 4 steps, have 1 step. GAP processes calculations above 768 channels with 3 × 3 dimensions as 1 × 1 @ 768, which is transformed as 16 × 16 @ 3

The Convolution Network identifies road areas, but does not guarantee intersections. Similarly, in shallow areas of RGB-D data, it does not guarantee compressed distances and therefore may generate irregular depth lines of RGB-D data. Therefore, the post-processing step is used to reduce unclassified areas in order to improve the presentation of existing objects.

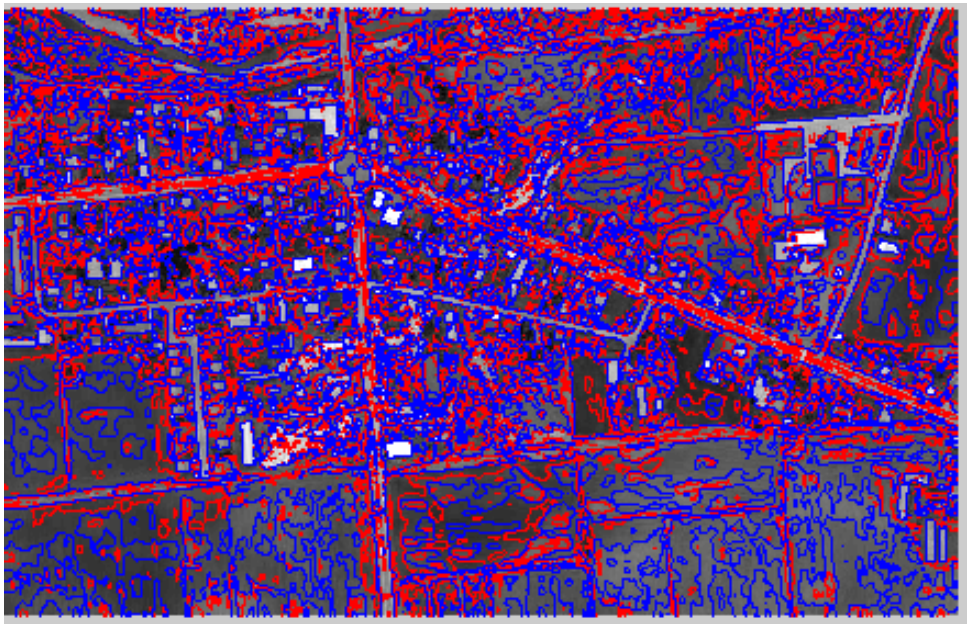


Figure 3: Features of normalized images

Data collection

MATLAB software has been used to implement the proposed method. The satellite image is for the affected area. These images are RGB type, with a scale of 1.25000 and a resolution of 0.1 m / pixel. The tragedy before September 12, 2006, the day of the shooting, we used satellite imagery on the site[1] to evaluate the proposed method.

[1]See website address: https://project.inria.fr/aerialimagelabeling/

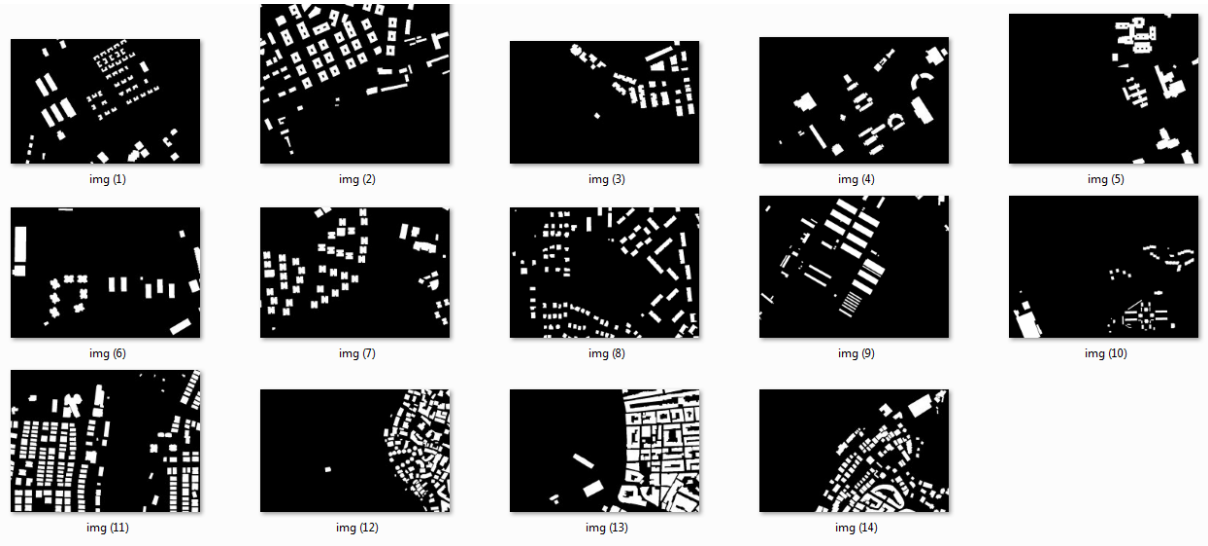


Figure 4: Test images

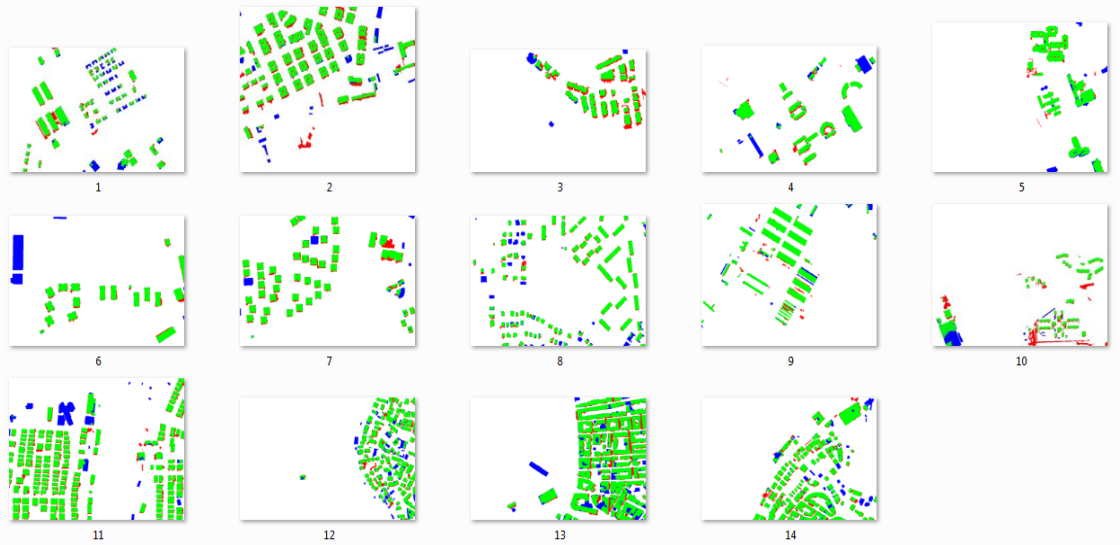


Figure 5: Output images

Table (1) compares the PSNR parameter for noise cancellation using the application of the mean filter as a basic method in noise cancellation for a small sample book image. The method of removing the proposed noise has given a better answer for different variations for the noise, and the resulting image is most similar to the original image without noise. It can also be said that the result obtained in the variance of higher noise is better than the result of lower variance. The impact of the business network on improving noise reduction has also been identified. It is clear that the higher the levels of decomposition, the better the noise reduction.

Table 1. Comparison of the PSNR parameter of the proposed book method with the basic methods

Noise Variance	Noisy Image	Intermediate Filter	Suggested Method
10	17.0709	17.0769	23.6443
15	15.0080	17.0693	23.6986
20	13.1470	17.1805	23.7197
25	11.5860	17.3495	23.8682
30	10.2817	17.3307	24.0535
35	9.1546	17.8707	24.2229
40	8.1720	18.2079	24.5380

The results of the second experiment

Table (2) compares the PSNR parameter for basic methods for large-scale image art. As can be seen from the results, the method of removing the proposed noise has given a better answer for different variations for the noise, and the resulting image is most similar to the original image without noise.

Table 2. Comparison of PSNR parameter proposed method with previous methods for art image

Noise Variance	Noisy Image	Intermediate Filter	Suggested Method
10	17.0171	18.9118	24.2243
15	14.9576	18.9551	24.2540

20	13.1315	19.0780	24.4580
25	11.5538	19.4983	24.7178
30	10.2495	19.4855	24.9389
35	9.1834	19.9690	25.3480
40	8.2649	20.5488	25.9927

Criteria for evaluating results

In this study, in order to better evaluate the results of the proposed method in this research, the results of the disturbance matrix for images were defined. In the discussion of artificial intelligence, this table is used to determine the value of evaluation indicators such as accuracy and precision.

Allergy

Sensitivity criterion means a ratio of positive items that the test marks correctly as positive. Here is a ratio of features that the algorithm correctly identifies as properties that belong to residential areas and structures. Mathematically, sensitivity is the result of dividing real positives into sums of real positives and false negatives. Some cases are summarized below, each of which is described below.

$$\frac{\mathrm{TP}}{\mathrm{TP}\ +\ \mathrm{FN}} \qquad 1$$

Specificity criterion (TNR[1])

The criterion of correctness means the ratio of the negative cases that the experiment correctly marks as negative. Here is a ratio of features that the algorithm correctly identifies as features that do not belong to residential and structural areas. Mathematically speaking, the truth is the result of dividing real negatives by the sum of real negatives and false positives.

$$\frac{TN}{TN\ +\ FP} \qquad 2$$

True Negative Rate (ACC)

The ratio of the sensitivity criterion to the true negative rate is calculated as the accuracy criterion of the algorithm.

$$\frac{TP + TN}{TP + FP + FN\ +\ TN} \qquad 3$$

[1] True Negative Rate

Comparison of experiments

In this section, the general results of the experiments are reviewed and compared. As can be seen from the table, the enhanced convolutional neural networks accounted for the least error among all experiments in determining the depth of RGB-D data. To ensure accuracy, the final results and images must be reviewed and approved by an expert. However, to evaluate and evaluate the efficiency of the proposed algorithm, the following criteria such as sensitivity, Specificity and Accuracy can be assessed to estimate the correct percentage of the proposed methods. Tables 2 and 3 show the evaluation criteria for the proposed method.

Table 3: Numerical results of the proposed algorithm (threshold 0.6))

F-score	Recall	Precision	Image
86/230	86/263	84/125	Img1
88/454	88/454	83/456	Img2
87/639	87/639	86/230	Img3
85/026	85/026	88/454	Img4
87/571	87/571	87/639	Img5
90/329	90/329	85/026	Img6
88/452	88/452	87/571	Img7
89/103	89/103	90/329	Img8
87/025	87/025	88/452	Img9
86/965	86/965	89/103	Img10
87/411	87/411	87/025	Img11
86/297	86/297	86/965	Img12
88/138	88/138	87/411	Img13
86/230	86/230	86/297	Img14
88/454	88/454	88/138	**Average**

Table 3: Numerical results of the proposed algorithm (3.6)

F-score	Recall	Precision	Image
86/230	86/263	84/125	Img1
88/454	88/454	83/456	Img2
87/639	87/639	86/230	Img3
85/026	85/026	88/454	Img4
87/571	87/571	87/639	Img5
90/329	90/329	85/026	Img6
88/452	88/452	87/571	Img7
89/103	89/103	90/329	Img8
87/025	87/025	88/452	Img9
86/965	86/965	89/103	Img10
87/411	87/411	87/025	Img11
86/297	86/297	86/965	Img12
88/138	88/138	87/411	Img13
86/230	86/230	86/297	Img14

88/454	88/454	88/138	**Average**

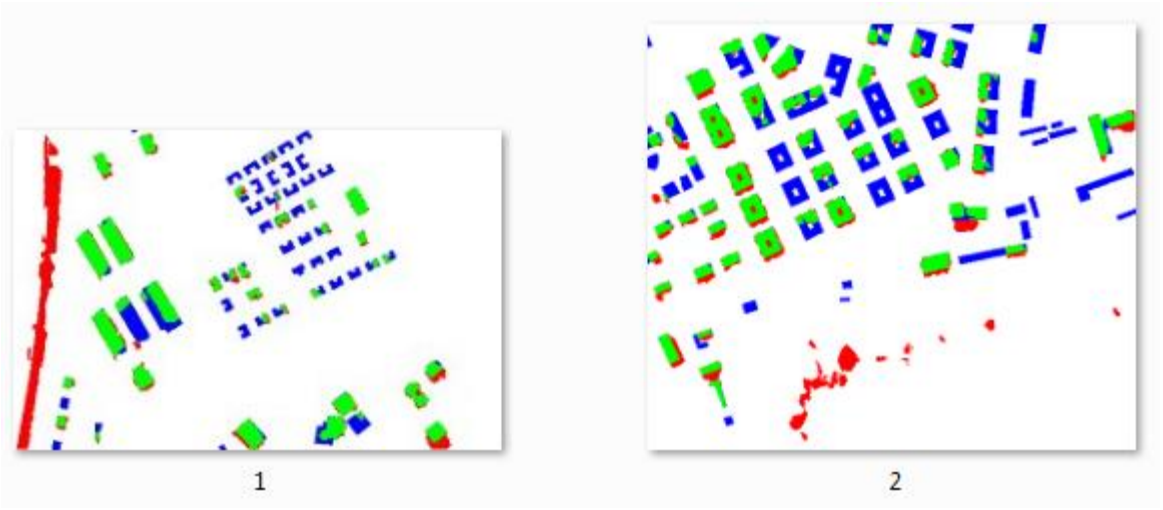
Table 3: Numerical results of the proposed algorithm

F-score	Recall	Precision	Image
86/230	86/263	84/125	Img1
88/454	88/454	83/456	Img2
87/639	87/639	86/230	Img3
85/026	85/026	88/454	Img4
87/571	87/571	87/639	Img5
90/329	90/329	85/026	Img6
88/452	88/452	87/571	Img7
89/103	89/103	90/329	Img8
87/025	87/025	88/452	Img9
86/965	86/965	89/103	Img10
87/411	87/411	87/025	Img11
86/297	86/297	86/965	Img12
88/138	88/138	87/411	Img13
86/230	86/230	86/297	Img14
88/454	88/454	88/138	**Average**

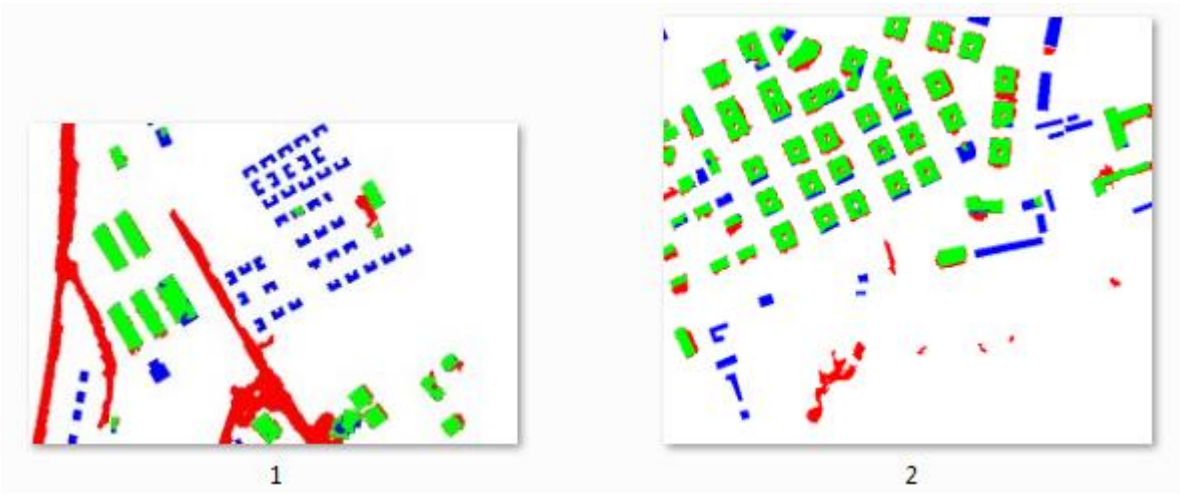
Comparison with other methods

In this section, to measure the performance of the proposed method in determining the depth of RGB-D data, it is compared with several methods (Figure 6-d). The output of these methods is the results of method [6] in 2019, in which the method of depth reduction of RGB-D data of independent remote sensing images is presented, which extracts a group of enhanced CNNs. The parameter-free method requires two parameters in the first step for post-processing. One of the advantages of this method is flexibility in the number of spectral bands. One of the limitations of this method and any method with supervision is training time. Research pathways in this regard include the automation of post-processing steps using improved integration criteria, which in turn leads to the full automation of the RGB-D data recovery mechanism (Figure 6-A). The method proposed in [10] suggests a multi-layered fusion model to restore depth from comparative RGB-D data and determine changes in remote sensing images in which path analysis or direct comparison is not feasible. This method uses uncontrolled or somewhat monitored clustering for a series of images with similarity measurements and subsequent resuscitation of multi-layered Markov random RGB-D data. After resetting the depth of the RGB-D data of each layer separately, the changes between each tagged map are detected. In this study, the advantage of the proposed method is numerically validated on a series of remote sensing images with real-world data. Of course, it should be noted that this issue requires more research in the field of higher quality images and the dimensions of feature vectors. Figure (6-b). Method 13 also identifies the depth of RGB-D data in satellite imagery. These images are processed using clustering techniques using color features to remove areas with vegetation and shadows that may have a negative effect on algorithms. HSV[1] is then presented for the image used, and the model presented is used to extract depth from RGB-D data. (Figure 6-c).

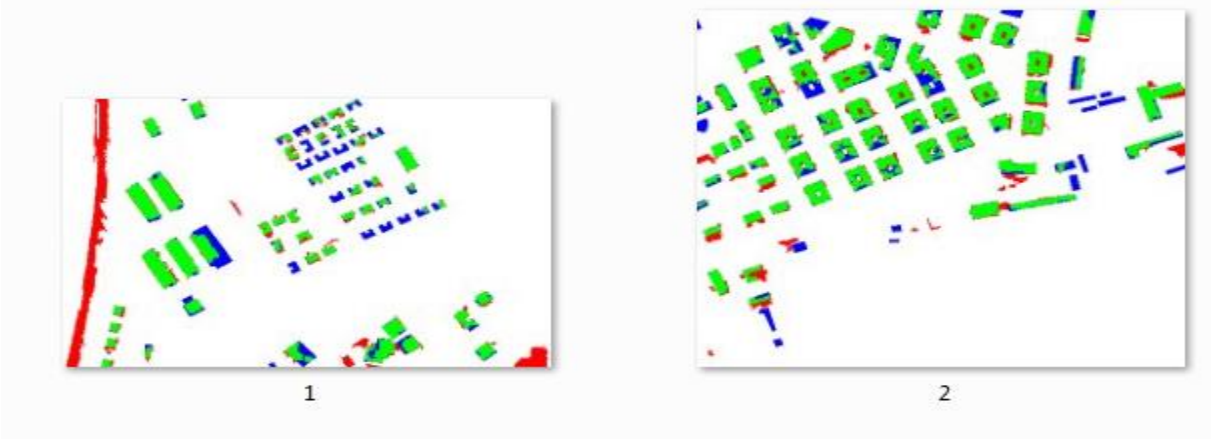
[1] Hue Saturation Value



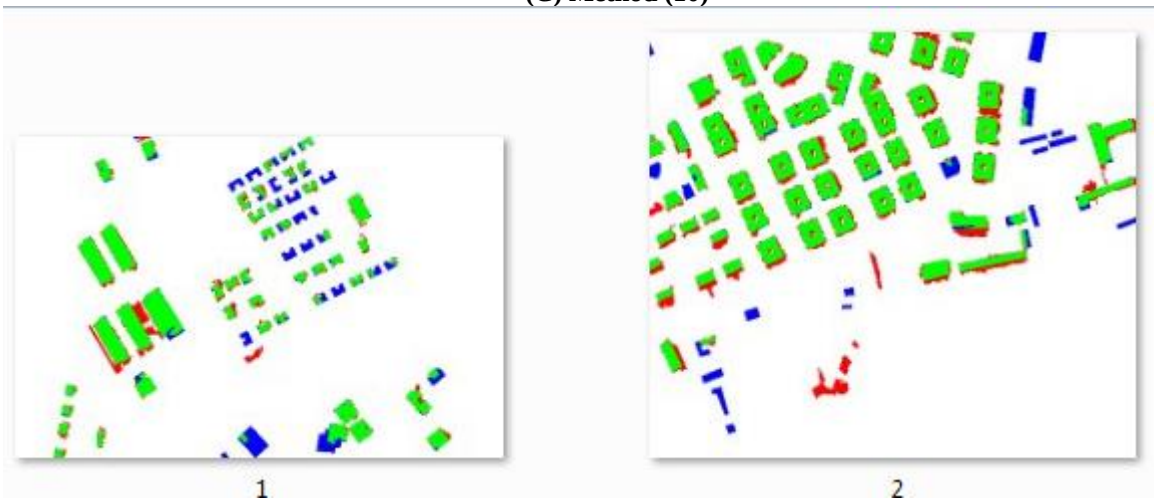
(A) Method (6)



(B) Method (13)



(C) Method (10)



(D) The proposed method

Figure 3: Results from different methods

So far, various methods and techniques have been developed to restore the depth of the data, each of which has been able to solve this problem to some extent.Based on the results of the table, the proposed method has been able to perform better than similar methods in terms of data recovery speed.

Table 4: compares the proposed method with other methods

Accuracy	Method
0.8701	**Method (10)**
0.8695	**Method (13)**
0.8315	**Method (6)**
0.8832	**The first proposed method (3)**

Conclusion

These methods maintain the boundaries in the image or texture, but are not very successful in reducing noise and cannot uniformly reduce image volume. In frequency methods, it is tried to model the destruction function of the imaging device and compensate for the destruction by using the model parameters. But usually for a device, the point noise destruction model is not constant and the noise distribution parameters in different tissues are different. Therefore, using a fixed model for all cuts of one volume does not have a suitable answer. In this paper, while reviewing a number of data recovery methods and comparing their advantages and disadvantages, an algorithm for image enhancement was presented, which both reduces point noise and does not damage image detail. In this paper, a method of data depth reduction based on metric learning method in reducing point noise in different images is proposed. In order to be more accurate in retrieving depth from data, noise variance is also calculated for each separate scale. In this way, our method becomes more sensitive to noise detection. The quantitative and qualitative results obtained from the implementation and calculation of the PSNR parameter of this method show that the proposed method of this paper has given a good answer compared to previous methods for noise elimination and has performed better in maintaining sharp corners and sharp features.

References

G. C. S. Gabarda, "Blind image quality assessment through anisotropy," *Journal of the Optical Society of America A*, vol. 24, no. 12, p. B42–B51, 2007.

M. B. A. H. I. W.J. Williams, "Uncertainty, information, andtime–frequency distributions," in *San Diego,'91, International Society for Opticsand Photonics*, San Diego, CA, 1991.

K. Z. D. T. Y. Y. X. G. W. Lu, "No-reference image quality assessmentin contourlet domain," *Neurocomputing*, vol. 73, no. 4, p. 784–794, 2010.

N. K. I. H. H. Liu, "A no-reference metric for perceived ring-ing artifacts in images," *IEEE Trans. Circ. Syst. Video Technol.*, vol. 20, no. 4, p. 529–539, 2010.

A. B. A.K. Moorthy, "Blind image quality assessment: from naturalscene statistics to perceptual quality," *IEEE Trans. Image Process*, vol. 20 , no. 12, p. 3350–3364, 2011.

A. B. C. C. M.A. Saad, "A DCT statistics-based blind image qualityindex," *IEEE Signal Processing Letters*, vol. 17, no. 6, pp. 583-586, 2010.

A. M. A. B. A. Mittal, "No-reference image quality assessment inthe spatial domain," *IEEE Trans. Image Process*, vol. 21, no. 12, p. 4695–4708, 2012.

A. B. X. W. C. Li, "Blind image quality assessment using a general regres-sion neural network," *IEEE TRANSACTIONS ON NEURAL NETWORKS*, vol. 22, no. 5, p. 793–799, 2011.

D. D. P. Ye, "No-reference image quality assessment using visual code-books," *IEEE Trans. Image Process*, vol. 21, no. 7, p. 3129–3138, 2011.

L. Y. P. L. Y. D. D. Kang, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

a. others, "On the Use of Deep Learning for Blind Image Quality Assessment," *CoRR*, vol. abs/1602.05531, 2016.

C. Q. L. M. J. G. a. R. H. Wei Zhang, "Learning Structure of Stereoscopic Image for No-Reference Quality Assessment with Convolutional Neural Network," *PatternRecognition*, vol. 59, p. 176–187, 2016.

D. M. K.-R. M. T. W. W. S. Sebastian Bosse, "Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment," *CoRR*, vol. abs/1612.01697, 2016.

H. S. L. C. a. A. B. Z.W., "LIVE Image Quality AssessmentDatabase Release 2".

D. C. E.C. Larson, "The CSIQ Image Database".

L. J. O. I. V. L. K. E. J. A. B. V. K. C. M. C. F. B. C.-C. J. K. N. Ponomarenko, "Color image database TID2013: Peculiarities and preliminary results," *Signal Processing: Image Communication*, vol. 30, p. 2015, 57-77.

X. Z. S. R. J. S. Kaiming He, "Identity Mappings in Deep Residual Networks," *arXiv:1603.05027* , 2016.

Mohammad Muntasir Rahman, Yanhao Tan, Jian Xue, Ling Shao, Ke Lu, 3D object detection: Learning 3D bounding boxes from scaled down 2D bounding boxes in RGB-D images, Information SciencesVolume 476February 2019Pages 147-15.

Ambroise Moreau, Matei Mancas, Thierry Dutoit, Depth prediction from 2D images: A taxonomy and an evaluation study, Image and Vision Computing, Volume 93, January 2020, Article 103825.

Siddharth Srivastava, Brejesh Lall, DeepPoint3D: Learning discriminative local descriptors using deep metric learning on 3D point clouds, Pattern Recognition Letters, Volume 1271 November 2019, Pages 27-36.