RESEARCHERID
THOMSON REUTERS

*[Singh,* 5(9): September 2018]
DOI- 10.5281/zenodo.1409653

ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

# SPARSITY-BASED IMAGE DENOISING VIA DEEP LEARNING AND STRUCTURAL CLUSTERING

**Mehak Khosla[1] & Ram Singh[*2]**
Department of Computer Science & Engineering, Punjabi University
Patiala, Punjab (INDIA)

## ABSTRACT

Image Denoising is still a major challenge in image processing. To restore noise free images deep learning are used nowadays. That are used to extract features from low level to high level and used many hidden layers. While there are two challenges in deep learning one is overfitting and second is regularization. Regularization include weight decay and sparsity. Inspired by the success of deep learning we combine the deep learning and structural clustering based sparse representation into one framework to enhance the algorithm. Our experiment result have shown which noise is better and give good result using different noise variance. The 12 generic natural images are taken and comparison table is made and shown which noise provide good result at different variance of noise.

***Keyword:*** *Image Denoising, Deep Learning, Sparse Representation, Clustering, Regularization*

## I. INTRODUCTION

Noise in an image is unwanted source that is added while image acquisition. Deep Learning is a part of machine learning that used multiple hidden layers for feature extraction and each layer use an input from previous layer. Deep learning work in supervised and unsupervised manner. In image restoration, the input consist of raw data that form matrix of pixels; the hidden layers are used to extract features from matrix of pixels. The first layer extract the pixel and edges; second layer organize the pixel and compress the pixel value and so on. The deep in deep learning refer the number of layers that are used in image. Deep Learning use substantial credit path (CAP) that are used to assign the value from input layer to output. CAP is a chain of transformation that describe connection between input and output layer. Deep learning method work on greedy layer-by-layer method and help to extract the useful features from image. Deep learning are generally described in terms of universal approximation theorem and probabilistic theorem. The universal approximation theorem work on principal of feed forward neural network in which single change in hidden layer result in continuous function. In 1989, George Cybenko first provide result for sigmoid activation function. The probabilistic theorem define the chance of ocurence of an event. The probabilistic theorem introduce dropout as well as regularization in neural network. Deep learning go through multiple layers of input hierarchy algorithm that applies nonlinear transformation and use it to create output. Number of iterations are used to reach the acceptable level of output. As number of iterations increased it lead to more accurate output. Deep Neural Network work on feed forward network in which values are assigned and data flow from input to hidden layers to output. There is no loop back. Deep network works on virtual neurons in which values are assigned automatically to make connection between them. The weights and input values are multiplied and return an output between 0 and 1. If the values didn't match then values are changed accordingly to get more accurate result. Deep learning process include understanding the problem and predict whether it fit in deep or not. Identify the relevant data set and analyze them and then apply the algorithm. Train algorithm on large amount of labeled data set. And the test the data performance against the unlabeled data set. While deep learning provide very good result and accurately compute the values but there is two challenges in deep learning one is computational time and second is overfitting. DNN are prone to overfitting because of the added hidden layers that are used for feature extraction. Regularization method such as weight decay and sparsity are used to combat overfitting. Secondly, dropout method are used that omit values from hidden layers to exclude rare dependency. Cropping and rotation method are used to train the data set for overfitting. DNN may include various parameters such as size, weight, learning rate etc. for optimal space and computation time swapping of parameters are not good option. To overcome computation time batching is done to increase the speed. Sparse representation is associated with compressed sensing which was developed by Donho [1] . Candes et al. [2] demonstrate that original signal can be recovered by using small portion of Fourier Transformation. Compressed Sensing theory based upon three components : sparse representation, encoding measuring and reconstructing algorithm.

In this paper, we put forward a new image processing model called clustering-based sparse representation (CSR) that are based upon deep learning. In this model we combine local and nonlocal image processing signals into one structural model. We implement iterative regularization and re-weighted minimization [3] that help to control the number of iterations [4] and assign the weight for connected the nodes. We have use different types of noises at different standard variance and compared there PSNR values and analyze the result to show which noise give better result at different variance.

## II. CLUSTERING-BASED SPARSE REPRESENTATION MODEL

Clustering algorithm are used in data mining and pattern recognition problem. The goal of clustering is grouping set of data into clusters. The success of clustering depend on choice of similarity measure.
We use an data set of images X and a sparse coefficient $\alpha = \{\vec{\alpha}\}$ that are used to make connection and known as sparseland model [5] .
From image set X we extract patch $x_i$ at location i.

$$x_i = W_i X \quad (1)$$

Where $W_i$ denote rectangular windowing operator. When we extract patches from image the overlapping is done that is redundant. And can be obtained from least-square solution

$$X = (\sum_i W_i^T W_i)^{-1} (\sum_i W_i^T x_i), \qquad (2)$$

To reduce overlapping of patches we do averaging of patches. To extract useful information from each patch belong to sparse co-efficient $\{\vec{\alpha}\}$ and information is given by dictionary $\Phi$ by

$$x_i = \Phi \vec{\alpha}_{i,} \qquad (3)$$

By adding Eq. (3) into Eq. (2), we obtain

$$X = D\vec{\alpha} = (\sum_i W_i^T W_i)^{-1} (\sum_i W_i^T \Phi \vec{\alpha}_i), \quad (4)$$

Where D is the operator dual to W . One variational problem under image denoising

$$\vec{\alpha} = \arg\min \frac{1}{2} \|Y - D\vec{\alpha}\|_2^2 + \lambda \|\vec{\alpha}\|_{1,} \qquad (5)$$

Where $Y$ contain distorted image signal and $\lambda$ stands for standard Lagrangian multiplier. The key motivation is sparse coefficient $\vec{\alpha}$ are not randomly distributed. Various algorithms are developed to solve the above convex optimization problem [6].

The main idea behind is sparse coefficient $\vec{\alpha}_i$ are not randomly distributed. The higher sparsity is achieved by location related constraint. To solve intensity and location uncertainty problem by bilateral filtering originally proposed in [7]. To establish a connection between data clustering and sparse representation is not possible because they are inspect at different levels. To enhance the performance of non local with sparsity, we study cost function and understand how they work.

$$(\vec{\alpha},\vec{\mu}) = \arg\min \frac{1}{2}\|Y - D\vec{\alpha}\|_2^2 + \lambda_1\|\vec{\alpha}\|_1 + \lambda_2\sum_{k=1}^{k}\sum_{i\in C_k}\|\Phi\vec{\alpha}_1 - \vec{\mu}_k\|_2^2 \qquad (6)$$

Where $\vec{\mu}_k$ is centroid of k-th cluster. The new clustering based regularization is that weight co-efficient $\vec{\alpha}$ are re-encoded with $\vec{\mu}_k$.

To make the regularization term more significant, we rewrite Eq.(6) as

$$(\vec{\alpha},\vec{\beta}) = \arg\min \frac{1}{2}\|Y - D\vec{\alpha}\|_2^2 + \lambda_1\|\vec{\alpha}\|_1 + \lambda_2\sum_{K=1}^{K}\sum_{i\in C_k}\|\Phi\vec{\alpha}_1 - \vec{\mu}_k\|_2^2 \qquad (7)$$

Where $\vec{\mu}_k = \Phi\vec{\beta}_k$ that is all centroid of vector are represented with same dictionary. We have unitary property of dictionary learning $\Phi$, as $\left\|\Phi\vec{\alpha}_i - \Phi\vec{\beta}_k\right\|_2^2 = \left\|\vec{\alpha}_i - \vec{\beta}_k\right\|_2^2$. Therefore, Eq. (6) overcome the optimization problem.

$$(\vec{\alpha},\vec{\beta}) = \arg\min \frac{1}{2}\|Y - D\vec{\alpha}\|_2^2 + \lambda_1\|\vec{\alpha}\|_1 + \lambda_2\sum_{K=1}^{K}\sum_{i\in C_k}\left\|\vec{\alpha}_i - \vec{\beta}_k\right\|_1 \qquad (8)$$

CSR model help to understand sparsity by combing deep learning and structural clustering into one framework.

## III. ITERATIVE REWEIGHTED AND REGULARIZATION MINIMIZATION

To solve optimization problem of Eq. (8) we use an iterative algorithm that are used to update $\vec{\alpha}$ and $\vec{\beta}$ via surrogate function [8]. To update $\vec{\alpha}$ by fixing $\vec{\beta}$, we use iterative shrinkage algorithm i.e

$$\alpha_j^{(i+1)} = \begin{cases} S_{\tau_1,\tau_2}(v_j^{(i)}) & \beta_j \geq 0 \\ -S_{\tau_1,\tau_2}(-v_j^{(i)} & \beta_j < 0 \end{cases} \qquad (9)$$

Where

$$v^{(i)} = \frac{1}{c}D^T(x - D\alpha^{(i)}) + \alpha^{(i)} \qquad (10)$$

And $\tau_1 = \frac{\lambda_1}{c}$ and $\tau_2 = \frac{\lambda_2}{c}$ where c is an auxiliary parameter, i denote iteration number and j denote j-th entry in vector. From the above equation we conclude that to overcome the problem of optimization we use iterative shrinkage and can be used in two regularization parameter of local and nonlocal sparsity respectively [8] .

To enhance the performance of an algorithm, we need to do some changes to improve CSR algorithm and its associated optimization algorithm. To adjust two regularization parameter $\tau_{1,}\tau_2$ we have variational image restoration [9] and optimization [9]. In [10], it was shown that signal-to-noise ratio is inversely proportional to regularization parameter $\lambda$, in [9] it is mention that for compressed sensing, signal magnitude $|x|_1$ is inversely proportional to new weights. Therefore, to update $\tau_1,\tau_2$:

$$\tau_1 = c_1\frac{\sigma_w^2}{\sigma_w}, \tau_2 = c_2\frac{\sigma_w^2}{\sigma_\gamma}, \qquad (11)$$

Where $\sigma_w^2$ is different level of noise that are added in the image, $\vec{\gamma} = \vec{\alpha} - \vec{\beta}$ and $c_1, c_2$ are predefined constant. Secondly, inspired by [10] work, we do more changes to recovered image by

$$X^{(i+1)} = \tilde{S}(1-\delta)X^{(i)} + \delta Y), \qquad (12)$$

Where $\tilde{S} = D \circ S \circ R$ projection into regularization set and

$$(1-\delta)X^{(i)} + \delta Y = X^{(i)} + \delta(Y - X^{(i)}), \qquad (13)$$

Is iterative regularization. The RHS of Eq. (13) is decline Land-weber operator where the blurness in the image is reduce to identity operator and $\delta$ is small positive number controlling the noise. We terminate the algorithm after three iterations. A complete description of CSR algorithm are given below:

**ALGORITHM 1 : IMAGE DENOISING VIA CSR**

1) Initialization: $\hat{X} = Y$;
2) Outer loop (dictionary learning): for i = 1,2,3…….,I
   - Update $\Phi$ via k-means and PCA;
3) Inner loop (Structured Clustering): for j = 1,2,3….,J
   - Iterative regularization: $\tilde{X} = \hat{X} + \delta(Y - \hat{X})$;
4) Regularization parameter update : obtain new estimate of $\tau_1, \tau_2$ via Eq.(11);
   -Centroid estimate update: obtain new estimate of $\vec{\beta}_k$ via kNN clustering;
   -Image estimate update: obtain new estimate of X by
   $\tilde{X} = D \circ S \circ R\tilde{X}$;

## IV. IMAGE DENOISING EXPERIMENT

We have put forward CSR denoising algorithm and implement under MATLAB (source code taken from [11]. The parameters that are used in experiment are: block-size B = 7 , λ = 0.03, deep-size k = 64, and I = J = 3. In this paper, we combine deep learning and structural clustering into one unified framework to enhance the performance of an algorithm. Deep Learning is implemented by k-means and PCA while clustering is implemented by BM3D. When image of highly self repeating, deep learning play important role than structural clustering and provide good result but combing together improve the performance and enhance the algorithm.

We have also compare the CSR algorithm and different type of denoising noise in literature at different level for a collection of 12 images. In deep learning and structural clustering to test the data 12 image set is taken and different noise variance is added in images to check which noise provide better result. In literature, additive Gaussian Noise is taken and there PSNR value of 12 images are taken at different point. In this paper, we take different types of noises like speckle noise, salt and pepper noise and Poisson noise. And there PSNR value at different noise variance are studied and compared with additive Gaussian noise and comparison Table-1 is made to show which noise give better result by using deep learning and structural clustering technique. Fig. 1 show denoising performance of Lena image using σ = 25. Fig. 2 show denoising performance of Boat image at σ = 15. Fig: 3 show denoising performance of pepper image at σ = 10.

ISSN 2348 – 8034
Impact Factor- 5.070

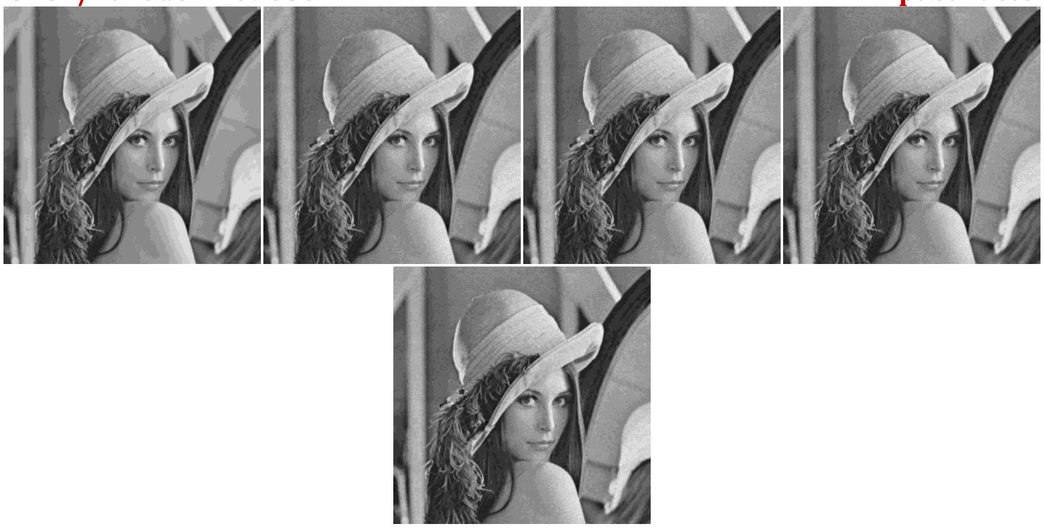

***Figure 1: Denoising performance for Lena image: a) original image b) Additive Gaussian Noise (PSNR = 32.08) c) Speckle Noise (PSNR= 32.56) d) Poisson Noise (PSNR = 32.41) e) Salt and pepper Noise (PSNR = 32.56) (σ = 25)***

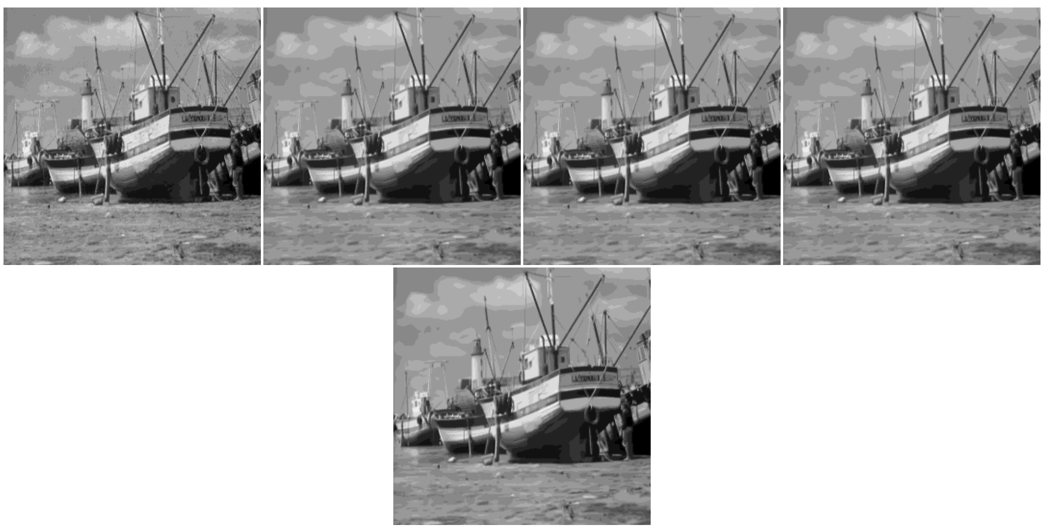

***Figure 2: Denoising performance for Boat image: a) original image b) Additive Gaussian Noise (PSNR = 32.15) c) Speckle Noise (PSNR = 31.80) d) Poisson Noise (PSNR = 32.06) e) Salt and pepper Noise (PSNR = 31.81) ( σ = 15)***

ISSN 2348 – 8034
Impact Factor- 5.070

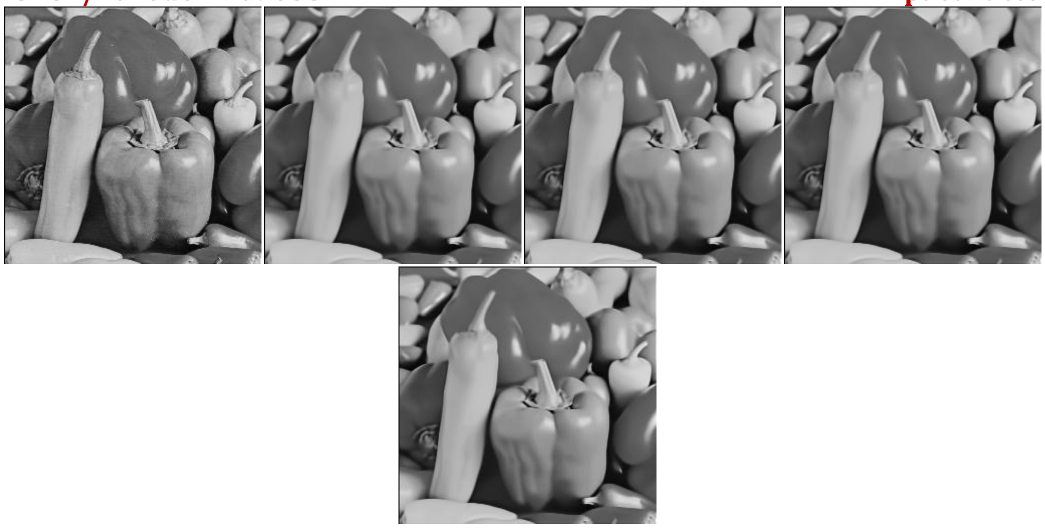

*Figure 3: Denoising performance for Pepper image: a) original image b) Additive Gaussian Noise (PSNR = 34.72 ) c) Speckle Noise (PSNR = 34.58 ) d) Poisson Noise (PSNR = 33.35) e) Salt and pepper Noise (PSNR = 34.74) ( σ = 10).*

*Table 1: The PSNR(dB) results for different denoising noise.In each cell, The result of four denoising noise are reported. Top Left: CSR(Additive Gaussia Noise), Top Right: Speckle Noise, Bottom Left: Poisson Noise, Bottom Right: Salt and Pepper Noise*

| σ | 5 | | 10 | | 15 | | 20 | | 25 | | 30 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lena | 38.74 | 38.63 | 35.94 | 35.86 | 34.29 | 34.29 | 33.07 | 33.24 | 32.08 | 32.56 | 31.28 | 32.05 |
| | 29.29 | 36.46 | 34.52 | 36.00 | 34.31 | 34.31 | 33.19 | 33.27 | 32.41 | 32.56 | 31.86 | 32.03 |
| Monarc-h | 38.43 | 40.01 | 34.49 | 34.49 | 32.25 | 31.88 | 30.71 | 30.35 | 29.53 | 29.45 | 28.56 | 28.83 |
| | 29.52 | 36.33 | 33.48 | 34.76 | 32.28 | 31.88 | 30.47 | 30.36 | 29.41 | 29.43 | 28.74 | 28.81 |
| Barbara | 38.43 | 38.95 | 35.10 | 34.77 | 33.17 | 32.75 | 31.78 | 31.41 | 30.66 | 30.56 | 29.77 | 30.04 |
| | 29.30 | 35.96 | 33.76 | 35.07 | 33.11 | 32.77 | 31.50 | 31.40 | 30.55 | 30.58 | 29.92 | 30.02 |
| Boat | 37.31 | 37.07 | 33.92 | 33.57 | 32.15 | 31.80 | 30.89 | 30.60 | 29.92 | 29.82 | 29.11 | 29.24 |
| | 28.68 | 35.52 | 32.91 | 33.82 | 32.06 | 31.81 | 30.62 | 30.59 | 29.72 | 29.83 | 29.11 | 29.22 |
| Camera man | 38.29 | 39.42 | 34.13 | 33.15 | 31.91 | 31.15 | 30.51 | 29.97 | 29.51 | 29.12 | 28.70 | 28.51 |
| | 29.06 | 36.45 | 33.37 | 33.67 | 31.42 | 31.17 | 30.08 | 29.98 | 29.15 | 29.15 | 28.48 | 28.52 |
| Couple | 37.49 | 37.61 | 34.02 | 33.56 | 32.10 | 31.67 | 30.75 | 30.42 | 29.70 | 29.63 | 28.84 | 29.08 |
| | 29.08 | 35.92 | 33.08 | 33.85 | 31.93 | 31.64 | 30.45 | 30.41 | 29.56 | 29.64 | 28.96 | 29.07 |
| Finger Print | 36.85 | 38.39 | 32.70 | 32.87 | 30.47 | 30.26 | 28.97 | 28.57 | 27.84 | 27.56 | 26.95 | 26.74 |

ISSN 2348 – 8034
Impact Factor- 5.070

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 27.83 | 33.82 | 31.05 | 33.11 | 30.56 | 30.26 | 28.77 | 28.58 | 27.63 | 27.56 | 26.80 | 26.74 |
| Hill | 37.12 | 36.97 | 33.66 | 32.81 | 31.88 | 31.27 | 30.73 | 30.35 | 29.85 | 29.78 | 29.14 | 29.41 |
| | 29.56 | 35.47 | 32.83 | 33.10 | 31.50 | 31.27 | 30.32 | 30.34 | 29.66 | 29.79 | 29.23 | 29.40 |
| House | 39.98 | 39.43 | 36.88 | 36.60 | 35.11 | 35.27 | 33.92 | 34.56 | 32.99 | 33.96 | 32.21 | 33.48 |
| | 28.74 | 36.09 | 34.45 | 36.71 | 35.18 | 35.28 | 34.19 | 34.53 | 33.57 | 33.97 | 33.12 | 33.46 |
| Man | 37.80 | 38.16 | 33.96 | 33.19 | 31.91 | 31.17 | 30.56 | 29.99 | 29.56 | 29.32 | 28.81 | 28.88 |
| | 29.56 | 36.24 | 33.12 | 33.51 | 31.49 | 31.17 | 30.01 | 29.98 | 29.23 | 29.32 | 28.75 | 28.88 |
| Pepper | 38.09 | 38.43 | 34.72 | 34.58 | 32.75 | 32.61 | 31.31 | 31.23 | 30.23 | 30.39 | 29.31 | 29.76 |
| | 29.00 | 35.72 | 33.35 | 34.74 | 32.85 | 32.56 | 31.30 | 31.23 | 30.34 | 30.37 | 29.69 | 29.74 |
| Straw | 35.89 | 38.90 | 31.51 | 32.20 | 29.14 | 29.24 | 27.50 | 27.23 | 26.21 | 25.84 | 25.16 | 24.90 |
| | 26.87 | 32.38 | 29.49 | 32.33 | 29.41 | 29.22 | 27.43 | 27.22 | 25.96 | 25.86 | 24.97 | 24.92 |
| Average | 37.85 | 38.49 | 34.23 | 31.09 | 32.23 | 31.87 | 30.84 | 30.65 | 29.79 | 29.86 | 28.95 | 29.24 |
| | 28.87 | 36.03 | 32.95 | 34.14 | 32.17 | 31.94 | 30.69 | 30.65 | 29.76 | 29.83 | 29.14 | 29.23 |

From the above table we conclude that when the noise in image is small there PSNR value is high and when the noise in image is high there PSNR value decrease. Fig. 4 and 5 show the graph of Additive Gaussian Noise and Salt and pepper Noise.

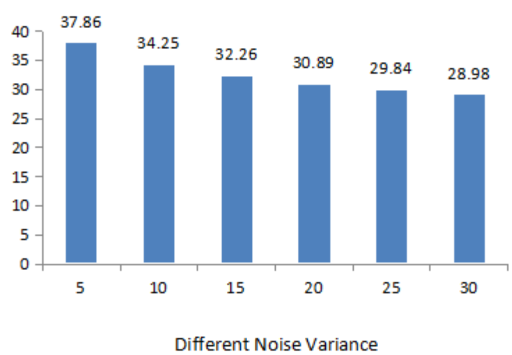


***Fig 4: Graph of Additive Gaussian Noise at different variance of Noise***

ISSN 2348 – 8034
Impact Factor- 5.070

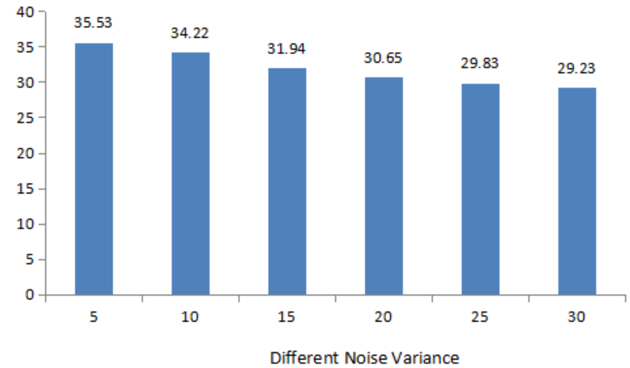


***Fig 5: Graph of Salt and pepper Noise at different variance of noise***.

Overall analysis is done to show which noise give best performance to enhance the performance of algorithm. Figure 6 show the result of noisy image when σ = 5 that show which noise give good result. Figure 7 show overall analysis result of different noise. From which we came to know which noise gives better result to enhance the performance of an algorithm.

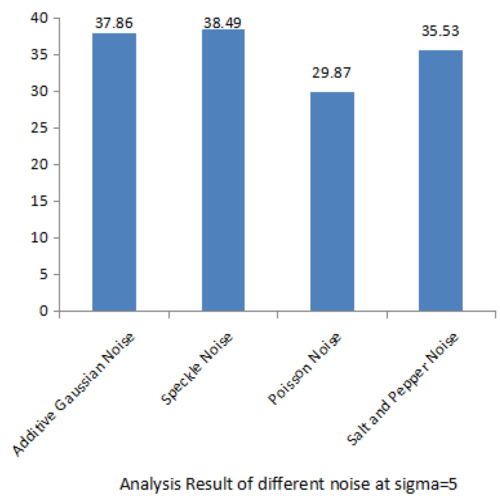


***Fig 6: Analysis result of different noise at σ = 5.***

[*Singh*, 5(9): September 2018]
DOI- 10.5281/zenodo.1409653
ISSN 2348 – 8034
Impact Factor- 5.070

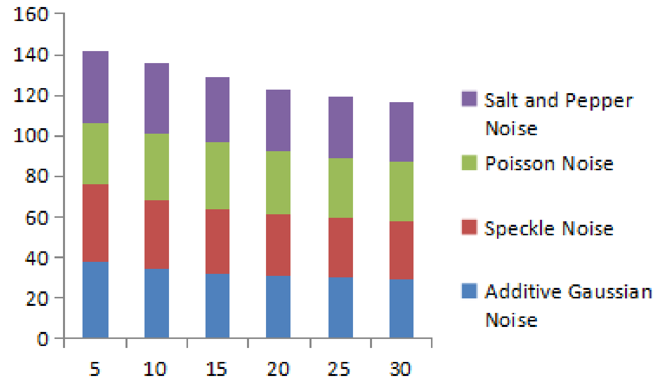


***Fig 7: Overall analysis result of different noise at different variance***

## V. CONCLUSION

It is an area to give new direction and help to understand the relation between deep learning and structural clustering. The collection of patches from natural image is a main challenge from a nonlinear and form a constellation; how to discover nonlinear using local geometry is a problem that has attracted much attention. Image denoising can be used under the framework of learning/reconstruction but unsupervised learning work with noisy data. From above we conclude that when noise in image increase there performance degrade but when noise in image is less we get better result. The main problem is when image contain too much noise then how to remove that noise to increase the performance and how to get better result.

## REFERENCES

1. *D. L. Donho, " Compressed sensing", IEEE Transaction on Information Theory, vol 52 no.4, pp 1289-1306, 2006*
2. *E. J. Candes, J.Romberg and T. Tao, " Robust Uncertainty Principles: Exact signal reconstruction from highly incomplete frequency information" IEEE Transaction on information Theory, vol 52 no 2, pp 489-509, 2006*
3. *E. Candes, M. Wakin and S. Boyd, " Enhancing sparsity by reweighted l1-miniimization," Journal of Fourier Analysis and Applications, vol 14 no 5, pp 877-905, 2008.*
4. *S. Osher, M. Burger, D. Goldfarb, J. Xu and W.Yin, " An iterative regularization method for total - variation based image restoration," Multiscale Modelling and Simulation, vol no 2, pp 460-489, 2005.*
5. *K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, " Image denoising by sparse 3-d transform-domain collaberative filtering," IEEE Trans.on Image Processing, vol 1 no 8, pp. 2080-2095, Aug 2008.*
6. *M. Zibulevsky and M. Elad, " L1-12 optimization in signal and image processing," IEEE Signal Processing Magazine, vol 27 no 3, pp 76-88, may 2010.*
7. *C. Tomasi and R. Manduchi, " Bilateral filtering and gray and color images," in ICCV, pp 839-846, 1998.*
8. *I. Daubechies, M. Defrise and C. De Mol, " An iterative thresholding algorithm for linear inverse problem with sparsity constraint", Communication on Pure and Applied Mathematics, vol 57 no 11, pp. 1413-1457,2004.*
9. *N. Glatsanos and A. Katsaggelos, "Methods for choosing the regularization parameter and estimating noise variance in image restoration and their relation," IEEE Transaction on Image Processing, vol 1 no 3, pp 322-336, Mar. 1992.*
10. *M. Zeler, D. Krishnan, G. Taylor and R. Fergue , "Deconvolution networks," in IEEE conf. On Computer Vision and pattern Recognition (CVPR), pp 2528-2535, 2010.*
11. *W.Dong, X. Li, L. Zhang, G. Shi , " Sparsity-based image denoising via dictionary learning and structural clustering," IEEE conference on Computer vision and Pattern Recognition ,pp 457-464, 2011*