International Journal of Engineering & Technology, 7 (2.33) (2018) 1018-1021

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Beyond a guassian denoiser: CNN for video denoising

Siyana E1 [1] *, Prof. Abid Hussain M [1]

[1] Department of Electronics And Communication TKM College of Engineering ,Kollam ,India
**Corresponding author E-mail: siyanaebrahim@gmail.com*

Abstract

Convolutional neural network are unique sort of neural network. They have so far effectively applied to video restoration errands. The proposed CNN is learned to exploit both spatial and temporal redundancy of video. We investigate the use of discriminative (conditional) model learning for video denoising. In this paper, take one wander forward by examining the improvement of feed forward denoising convolution-al neural frameworks (DnCNNs) to grasp the advance in profound design, and regularization procedure into video denoising. Specifically, residual learning and batch normalisation are utilized to quicken the training procedure and furthermore help the denoising execution.

Not exactly the same as the current conditional denoising models which never mention about the misalignment and for the most part train a specific model for Gaussian upheaval at a specific commotion level .The work have already been implemented for image but nowadays deep learning has great progress in computer vision that demand large amount of data. To the best of our knowledge, our method is proposed to extend the work in video denoising task like Gaussian denoising, video super-resolution and Video deblocking,which ultised the same method inorder to make good use of multiple frames based on CNN. The proposed model has the ability to manage Gaussian denoising with unknown commotion.

Keywords: *Video Denoising; Convolutional Neural Network; Residual Learning; Batch Normalization.*

1. Introduction

Video denoising can be described as the problem of removing noise from a noisy signal.Nowadays ,deep learning has made great progress in computer vision , image description. Main objective of video or multi frame denoising is to recuperate perfect video s from an uproarious perception r which takes after a video corruption model $r = s + h$. One typical presumption is that h is white Gaussian noise. From a Bayesian perspective, when the probability is known, the video earlier demonstrating will have a central impact in video denoising [1]. In course of recent decagon, different technique has abused for displaying frame priors, counting NSS models [2]–[6], BM3D [3], sparse models [7]–[9], gradient models [10] .

In spite of their good denoising peculiarity, the vast majority of this strategies regularly experience the ill effects of two noteworthy downsides. To begin with, those strategies by and large include an intricate enhancement issue in the checking stage, influencing the process time consuming [8]. In this way, the majority of the strategies can scarcely accomplish elite without giving up computational effectiveness. Second, the models when all is said in done are nonraised and include a few physically picked parameters, giving a few space to support denoising execution.

CNNs has effectively experienced on video information [11]. Training for recuperation reason remains a remarkable issue in light of the fact that the video quality necessities for the training database are high since the yield of the CNN is the veritable video as opposed to only a name. In this paper, rather than taking in a discriminativel model [3] with an express casing, to begin with process the video into outlines that regards outline denoising as plain conditional learning issue, i.e., confining the commotion from a noisy video by feed forward convolution neural (CNN). The reasons of using CNN are two-cover. In any case, CNN with deep architecture is intense in growing the limit and flexibility for abusing video characteristics. Second, noteworthy advances have been refined on regularization and learning systems for planning CNN, including batch standardization [12] and residual learning [13].

All above mention techniques can be embraced in CNN in order to accelerate the training process and enhance the denoising. As opposed to straightforwardly the outputting the denoised video $\hat{s}$, this model is intended to foresee the residual video, $\hat{h}$ that is the differentiation within the uproarious perception and the dormant orderly video. All things considered, the proposed DnCNN absolutely ousts the inherent orderly video by exercises in the concealed layers. The batch norm method be additionally acquainted with settle and upgrade the training execution of DnCNN. Incidentally residual learning [13] and batch normalisation can profit by one another, what's more, their combination is compelling in speed up the training, what's more, push the denoising execution.

The work have already been implemented for image which never mention about the misalignment. Nowadays deep learning has great progress in computer vision, pattern recongnition application demand large amount of data. A deep convolutional neural network is learned to exploit video spatial redundancy as well as global temporal redundancy of video.

The proposed model exhibits better visual quality as well a quantitative measure than state of the art video denoising method. To the best of our knowledge, our method is proposed to extend the work in video denoising task like Gaussian denoising , video super-resolution and Video deblocking. which ultised the same method inorder to make good use of multiple frames based on CNN. The procedure is to take a few continous noisy frames as input and output images where the noise has been reduced. Results shows that removing on additive white Gaussian noise as competitive with the current state o the art.

In this paper which plans towards outline an all the more convincing Gaussian denoiser and observe while h is the contrast within the ground truth high resolution video and the bicubic upsampling of the low resolution video, the video corruption display for denoising may changed over as video super resolution issue; comparably, the video deblocking issue will demonstrated by a similar video debasement show by taking h as the distinction between the first video and the compacted video. In this sense, video super resolution and video deblocking will dealt with as two extraordinary instances of a "general" video denoising issue, however in video super resolution and video deblocking the commotion h is entirely variant from AWGN. The DnCNN [1] can in like manner gain promising results while being connected with a couple of general Video denoising errands. What's more, it show the reasonability of planning just a solitary DnCNN show for three general video denoising task.

2. Proposed method

This model broaden it for dealing with a few general video denoising undertakings. By what's more, extensive, setting up a deep CNN observe for a specific task generally incorporates arrange architecture design and model learning from training information. For sort out system design , adjust the VGG arrange [14] to make it sensible for video denoising. For video denoising first convert the video into frames and process the frames i.e each frames convert into patches. Set the extend of convolutional channel to be 3 × 3. In this manner, the receptive field [2]of DnCNN with deep of k ought to be (2k+1)×(2k+1). Extending residual field [2] size can make utilization of the setting data in bigger edge district. For better execution and adequacy, one key issue in building design is to set a fitting Depth for DnCNN.

2.1. Network architecture

The network comprised 20 full convolutionl layers,with no pooling . The input to the DnCNN is a noisy observation r = s+h. The depth of network be k, network have three sorts of layers, appeared in Fig. 1. For first layer it perform Convolution and Rectifer linear unit to the necessary layer, channels of size 3 × 3 × c are used to deliver 64 highlight maps, and rectifer linear unit (ReLU, max(0, •)) are perform threshold operation inorder to ultised for non linearity. where c speak to no of frame channels, i.e., c = 1 indicate dim casing and c = 3 indicate colour outline. In second layer it perform Convolution,batch norm and ReLU: 64 channels to measure 3×3×64 . For third layer it perform Convolution: c channels of size 3 × 3 × 64 are used to recreate the yield. To aggregate up, model demonstrate has two fundamental highlights: the residual learning [13] definition is embraced to learn P(r), and batch normalisation [12] is consolidated to fast training as well as enhance the denoising execution.

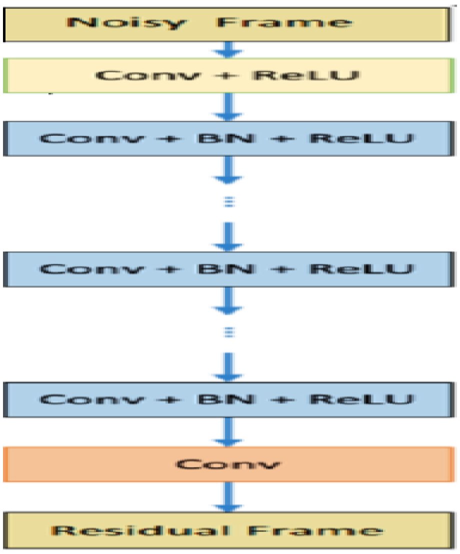

Fig. 1: Architecture of Proposed DNCNN.

By consolidating convolution with ReLU [15], DnCNN can step by step isolate frames from the uproarious observation through the shrouded layers[4]. In many visual processing experiment, it requires the resultant frame size should keep the same as the input one. This may the limit ancient rarities, straightforwardly cushion zeros preceding convolution to ensure that each component guide of the center layers has an indistinguishable size from the information casing and find that the basic zero cushioning system does not bring about any limit ancient rarities.

This great property is presumably ascribed to the intense capacity of the DnCNN. For DnCNN, receive the residual learning [13] definition to train a residual mapping P(r) ≈ h and after that s = r − P(r). For each frame, calculate averaged mean squared error between the desired residual frame and evaluated frame from uproarious information .

$$U\ (\Theta) = \frac{1}{2Q} \sum_{j}^{Q} ||P\ (r_{j;}\ \Theta) - (r_j - s_j)\ ||^2_F \qquad (1)$$

Can be taken as the loss function to learn the trainable parameters Θ [1]. Here {(rj, sj)} $_{j=}{}^{Q}$ represents noisy-clean training frame (patch) pairs. This strategy called residual learning is crucial to accelerate the training technique and enhances execution truth be told, it has been observed tentatively that training a CNN may be exceptionally direct when the coveted yield is generally the same as information. This is the circumstance of various restoration errand for instance, denoising [4] or super resolution. By defining the double objective of repeating the commotion preparing turns out to be much more viable, calculate the PSNR and SSIM values for each frame.

3. Experimental settings

To train model with known noise level, used freely avaliable video database which give high featured movies. The xylophone video sequence is utilized for training and testing .The video consist 30 scence from that it utilize 24 training and 56 for testing and calculate the normal PSNR and SSIM values from test outlines as an execution measure by taking noise levels, i.e., σ = 15, 25 and 50. furthermore, set the patch size 40x40 for training and allude to model for Gaussian for specfic noise level .Set the scope of the commotion levels as σ ∈ [0, 50], and the patch size as 50 × 50 to train the model for show for daze Gaussian denoising to prepare the model. Frames with a quality factor running from 5 to 99 utilizing the MATLAB encoder.

Network depth of 20 for three denoising task.. The loss function in Eqn.(1) be received to understand the residual mapping P(r) for foreseeing the residual h. Introduce the weights by the technique in [15] and utilize SGD [5] with weight decay of 0.0001, a momentum of 0.9 and a mini-batch size of 128 and train 50 epochs for models. The learning rate [6] was decayed exponentially from 1e−1 to 1e−4 for the 50 epochs.

All the tests are completed in the Matlab (R2013a) condition running on a PC. For Gaussian denoising [2], utilize the multiframe with clamor level of [0-50].and for video super resolution, prepared a solitary model for all the upscaling factors (i.e., 2 and3). For video deblocking, trained with quality factors 10 and 20 respectively.

4. Results and discussion

For Gaussian denoising, consider that the commotion range is known from fig (2) and from table 1, with noise level 15 and 25. In fig (2) upper left represent the noisy input, lower left represent the noise learnt from the noisy input, upper right represent output obtained and lower right represent the original frame and the table 1 shows the PSNR and SSIN values obtained for Gaussian denoiser.For video super resolution shown in fig (3) and table 2 with upscaling factor 2 and 3, this model can produce sharp edges and fine details for super resolution. In fig (3) upper left represent the noisy input, lower left represent the noise learnt from the noisy input, upper right represent output obtained and lower right represent the original frame and the table 2 shows the PSNR and SSIM values obtained for super resolution. Video deblocking shown in fig (4) and from table 3 with quality factor 10 and 20 respectively. This model recover the straight line, PSNR and SSIM values increase with hike in quality factor. In fig (4) upper left represent the noisy input, lower left represent the noise learnt from the noisy input, upper right represent output obtained and lower right represent the original frame and the table 3 shows the PSNR and SSIM values obtained for Video deblocking

Table 1: PSNR (Db) and SSIM for Gaussian Denoising with Noise Level 15and 25

Noise level	PSNR(dB)	SSIM
15	33.78	0.8891
25	31.65	0.8504

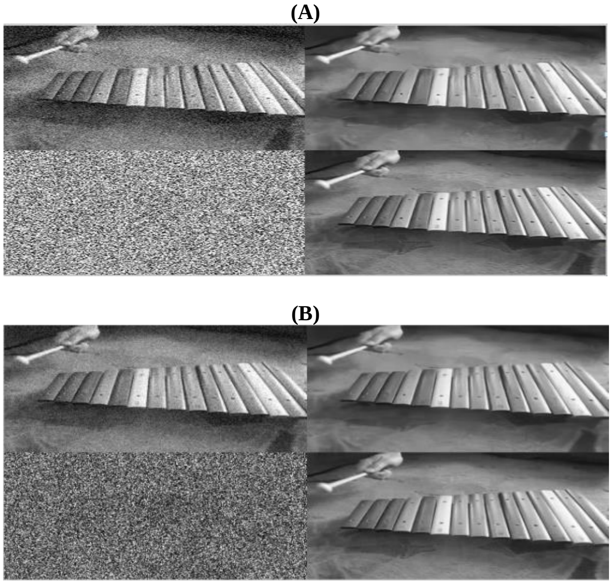
(A)
(B)

Fig. 2: Gaussian Denoiser, (A) Noise Level 15, (B) Noise Level

Table 2: PSNR (Db) and SSIM for Video Super Resolution with Upscaling Level 2and 3

Upscaling level	PSNR(dB)	SSIM
2	37.21	0.9537
3	33.06	0.9057

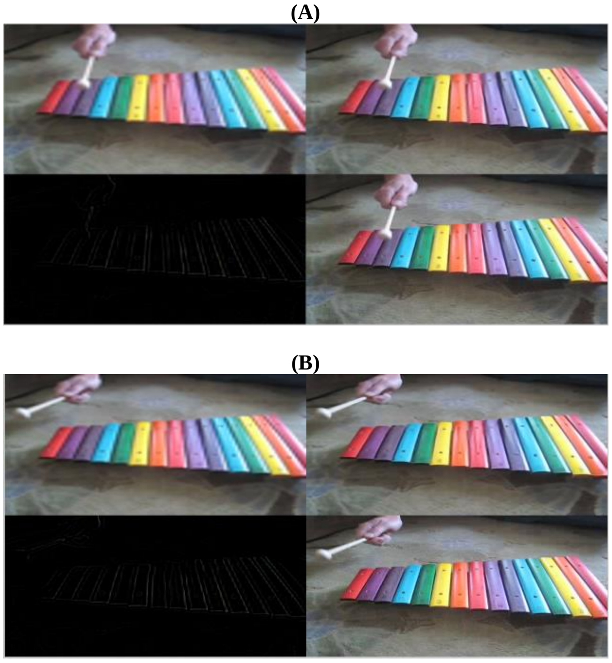
(A)
(B)

Fig. 3: Video Super Resolution, (A) Upscaling Factor 2, (B) Upscaling Factor 3.

Table 3: PSNR (Db) and SSIM for Video DE Blocking with Quality Factor Level 10 and 20

Quality factor	PSNR(dB)	SSIM
10	33.47	0.8697
20	34.06	0.9068

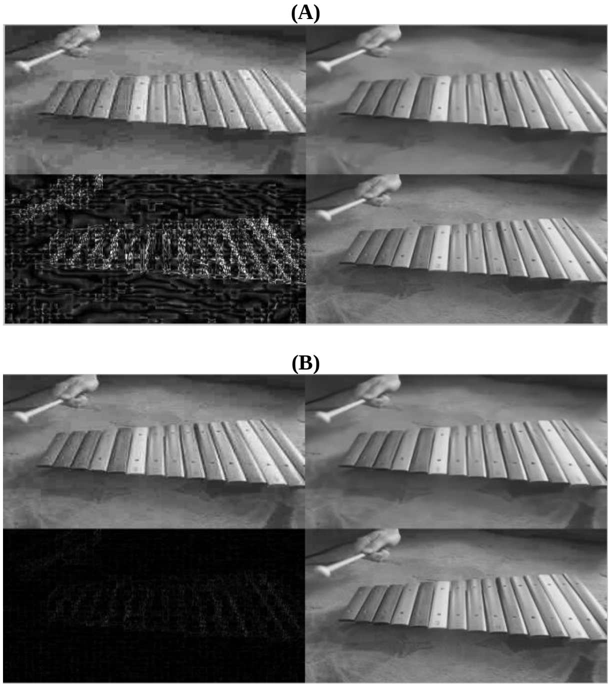
(A)
(B)

Fig. 4: Video De Blocking (A) Quality Factor 10, (B) Quality Factor 20.

5. Future scope

The combination of recurrent neural network and convolutional neural network can be used in video restoration to improve the performance. The CNNs are great at managing spatially related information while the RNNs are great at fleeting signs and accelerate the procedure.

6. Conclusions

In this paper we investigate the use of deep convolutional neural network to perform the video denoising, where residual learning be embraced to isolating commotion from uproarious perception. It exploits both spatial and temporal redundancy of video. The batch normalisation and residual learning are fused to accelerate the training technique and furthermore help the denoising execution. The conventional discriminative models which trained definite models for specfic commotion range, this model has the capability to manage with the outwardly hindered Gaussian denoising with obscure commotion range. It has the ability to handle the misalignment, video denoising make use of multiple frames and generated frames as real. Besides, it demonstrate practicality for train the model to deal with three general video denoising.

Acknowledgement

The author might want to thank prof. Abid Hussain M for giving appropriate guidelines throughout the work

References

[1] Kai Zhang, Wangmeng Zuo"Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," IEEE transactions on image processing, vol. 26, no. 7, july 2017.

[2] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2. Jun. 2005, pp. 60–65.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," IEEE Trans. Image Process., vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[4] A. Buades, B. Coll, and J.-M. Morel, "Nonlocal image and movie-denoising," Int. J. Comput. Vis., vol. 76, no. 2, pp. 123–139, Feb. 2008.

[5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in Proc. IEEE Int. Conf. Comput. Vis., Sep./Oct. 2009, pp. 2272–2279.

[6] J. Xu, L. Zhang, W. Zuo, D. Zhang, and X. Feng, "Patch group basednonlocal self-similarity prior learning for image denoising," in Proc. Int. Conf. Comput. Vis., Dec. 2015, pp. 244–252.

[7] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," IEEE Trans. Image Process., vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[8] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," IEEE Trans. Image Process., vol. 22, no. 4, pp. 1620–1630, Apr. 2013.

[9] Z. Zha et al. (2016). "Analyzing the group sparsity base on the rank minimization methods." [Online]. Available: https://arxiv.org/abs/1611.08983

[10] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Phys. D, Nonlinear Phenomena, vol. 60, nos. 1–4, pp. 259–268, 1992.

[11] Armin Kappelerand Seunghwan Yoo, "Video super resolution with convolutional neural networks," in IEEE Trans.on Computational Imaging ,vol.2,nos.2,June 2016

[12] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep a network training by reducing internal covariate shift," in Proc. Int. Conf.Mach.

[13] .K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*2015, pp. 1–14.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec.2015, pp. 1026–1034.