International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

Frame Prediction- Noise Removal using Denoising Autoencoders

Mrs. Manju D , Seetha M, Sammulal P

Abstract: ***In the current times, tasks like Object Detection, Object tracking, Gesture prediction, Video prediction in computer vision are being solved effectively with models of deep learning . Video frame prediction involves predicting the next few frames of a video given the previous frame or frames as input. Currently, the challenge in video frame prediction is that the predicted future frames are blurry. This paper focuses on the removal of noise from the predicted image using Denoising Autoencoders, solve the above-addressed issue. The proposed work, trains LSTM model which generates future frames by giving a sequence of input frames. The predicted output is given as an input to the Denoising Autoencoders which tries to remove the blurry predictions. Our approach is implemented on Moving MNIST Dataset. The result of our proposed method improved accuracy and is compared with the accuracy of Denoising Autoencoders, LSTM, and LSTM along with Denoising Autoencoders.***

Keywords : Denoising Autoencoders, Long short-Term memory (LSTM), Moving MNIST Dataset, Prediction.

I. INTRODUCTION

Computer Vision is a field, which includes fundamentals of images, object tracking, object detection, classification and video analysis. In computer vision, video analysis has become a challenging task and for solving the real life problems, Deep learning model plays a major role. Predicting the future frame has received much attention in unsupervised learning. The procedure is shown in Fig.1

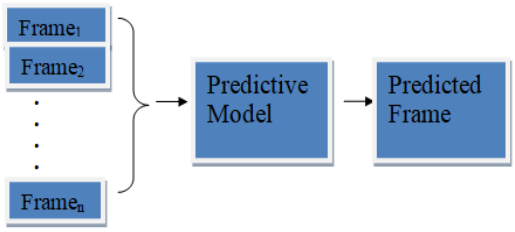

Fig. 1 Prediction of future frame is done by giving a sequence of input frames to the predictive model and in turn generates predicted frames.

Revised Manuscript Received on December 30, 2019.
***** Correspondence Author
Mrs. Manju D*, Assistant Professor, Dept.of CSE, GNITS, Hyderabad, India
Dr. Seetha M, Professor & HOD, Dept. of CSE, GNITS, Hyderabad, India
Dr. Sammulal P, Professor, Dept.of CSE, JNTUH CEJ, Hyderabad, India

This paper addresses the problem of blurriness in the predicted frames. For dealing with blurry predictions obtained after applying to any prediction model, a process by combining LSTM model with Denoising Autoencoders is proposed. Then compared the accuracy with the learning representation LSTM model with LSTM + Denoising Autoencoders.

Finally, Loss function has been used to calculate the inconsistency between the predicted value and the actual value. There are different loss functions in literature like Mean Square Error (MSE), L1 loss, Universal quality index, Adversial loss function etc. Loss function reduces the error in prediction. A value of '1'shows that the images are identical to each other or else the value is between '0' and '1'.

The paper is divided into the following sections, Section 2 overviews of existing methods, proposed work in Section 3, implementation results are shown in Section 4 and conclusion in Section 5.

II. LITERATURE SURVEY

In recent years, a huge amount of work has been carried out in video prediction and in the below mentioned research papers different authors have trained different models for predicting future frames and removal of blurriness from the predicted frames such as, 'Encoder and Decoder LSTM model' that worked on unlabeled videos and this model even helped for long-duration predicted frames [1]; Deals with blurry predictions obtained from the standard mean squared error function [2]; Suggested 'PGN network', which consists of convolutional Autoencoder with recurrent LSTM layer [3]; The solving of video generation task by using techniques to maintain temporal and spatial features [4]; The objective of the paper is to make, accurate prediction of future frames in complex and real-world scenes [5]; Proposed a 'two stage GANs', for predicting the future frames clearly. The loss function used here was Gradient Difference Loss (GDL). The experimentation was carried out on datasets like Sports-1M, UCF-101, KITTI [6]; An architecture called 'Fully Context-aware' was introdued,this architecture or model is able to have sharper predictions.The MD-LSTM blocks was used where the output of each are combined in context blending blocks[7]; Introduced 'Predictive Coding Network (PredNet)' where the system continuously makes predictions and compares them to the actual input. The error is then propagated so that the predictive model can be updated [8]; A two-stage frame work called 'Video Prediction via Selective Sampling (VPSS)' was introdued. Datasets like Human 3.6M, Moving MNIST and RobotPush were used for experimentation[9].

For capturing the spatial and temporal features in natural video's ,proposed an Encoder-Decoder CNN and convolutional LSTM model[10]; Proposed a method Deep Voxel flow that consists of a Fully-Convolutional Encoder-Decoder architecture containing with three convolutional and deconvolutional layers[12]; for predicting the video frame used Gated Recurrent units which confluence noise during training phase [13]; This paper focuses on pixel level video prediction that uses hierarchical approach[14]; Predicts the Pedestrian's behavior, by extracting the person appearance and pose features[16]; Proposed an SDC module for learning motion vector of objects in the frame prediction and a kernel for synthesizing the pixels[17,31]; Proposed a Meta-Learning an unsupervised learning rule for training networks[18]; Focused on action prediction sequences, for this proposed global-local temporal action prediction model[24,23,33].

Intuitively, all the last few predicted frames from different models suffer from blurriness.

III. PROPOSED METHOD

Long Short Term Memory (LSTM) network architecture has feedback connections that have designed to work with a single data point as well as a sequence of data points. LSTM network is one of the best-suited models for predicting the time series data and deals with vanishing gradient problem. LSTM unit consists of cells, where it stores the information. The cells make a decision about what information to store, read, write and delete which is controlled via gates. The three types of gates that are used are Input, Output and Forget gates. The two states that are transferred to the next layer are the Cell state and the Hidden state. It is shown in Fig 2 [25].

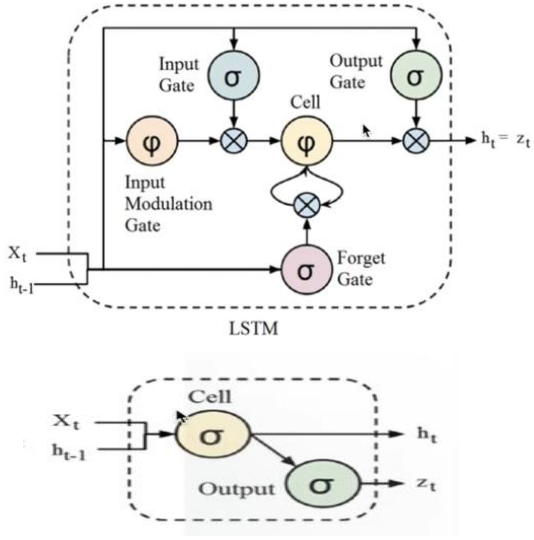

Fig. 2. LSTM model.

Forget gate: Makes decisions about what to remove from previous state h(t-1)

Equation: $f_t = \sigma (W_f.[h_{(t-1)}, x_t] + b_f)$ (1)

Where σ is sigmoid function.
W_f is Weight matrix
b_f is bias vector parameter
$h_{(t-1)}$ is previous state hidden vector
x_t is Input vector to LSTM

Input gate: What information need to be written to the cell.

Equation: $i_t = \sigma (W_i.[h_{(t-1)}, x_t] + b_i)$ (2)
$c_t = relu (W_c.[h_{(t-1)}, x_t] + b_c)$

Where c_t cell state vector
i_t is input gate activation vector

Output gate: what needs to be output to a cell.

Equation: $Ot = \sigma (Wo.[h(t-1), xt] + bo)$ (3)
$h_t = o_t * \tanh(c_t)$

where h_t is hidden state vector
o_t is an output gate activation vector
c_t is a cell state vector

Denoising Autoencoder is an extention of the basic autoencoder. Denoising autoencoder attempts to randomly corrupt input data.

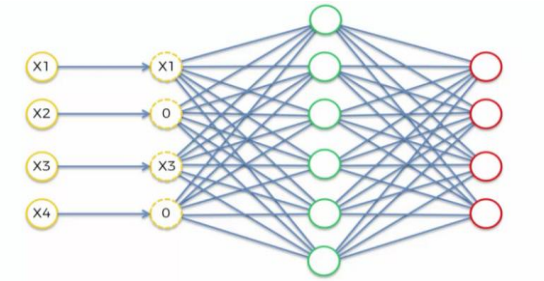

Fig. 3. Denoising Autoencoders.

In Denosing autoencoders, the input image is corrupted, by adding noise.Denoising auto encoder job is to recover the input . The input to the denoising autoencoder is the predicted frame and to this salt and pepper noise is added for the noise removal [31].

In Denoising autoencoder, the structure is

$$Y=[w\ b].[(p+r);1] \quad (4)$$

where r is the noise added to input.

A random value is generated for every patch

$$Y=f_{NL}[w\ b].[(p+r);1] \quad (5)$$

The process of the proposed method is shown below in the fig. 4, where the prediction model used is LSTM which generates the predicted frame. This predicted frame acts as an input to the Denoising Autoencoders which try to remove the blurriness in the predicted frame.

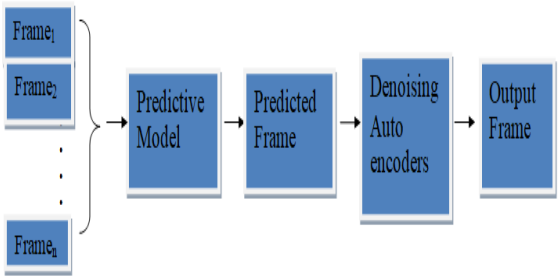

Fig. 4. Flow of the proposed method .

IV. RESULTS

The given method is implemented on Moving MNIST Dataset. The dataset consists of 10,000 sequences with length 20 showing 2 digits moving inside a 64X64 frame. To start the work we used front end as keras and back end as Tensorflow. For learning the environment, LSTM model is used , which inturn helps for predicting the next frames given the past frames as input and to evaluate the quality of predicted frame Mean Square Error is used. For the experimentation, epochs are used to measure the MSE of Training and Test dataset. On changing the different parameters like epochs and batch size above 250, best results can be achieved. Batch size tells how many frames are loaded. The input and output frame are divided by 255 to normalize all the data between '0' and '1'. The input frame is resized to 28X28, experimented with epoch value of 10 and batch size 150 and a comparison between the three models named Denoising Autoencoders, LSTM model, the combination of LSTM along with Denoising Autoencoders is shown in terms of loss and on number of epochs. However, The loss function used is MSE , it is the simplest of all the errors which is defined as error in the divergence of prediction from actual rating and the activation function used in Denoising Autoencoders is Relu. After calculating the loss, finally graph is plotted between number of epochs by loss.

TABLE-1: SUMMARIZES THE ACHIEVED ACCURACY VALUES FOR THE THREE MODELS BASED ON THE NUMBER OF EPOCHS.

Model	Total no. of Parameters	No.of Epochs	Accuracy
Denoising Autoencoders	49761	10	50%
LSTM Model	57828	10	31%
LSTM + Denoising Autoencoders	57828	10	33%

It is observed that Denoising Autoencoders have an accuracy of 50% when it is implemented directly on a dataset. Compared to the LSTM model the proposed method accuracy is increased.

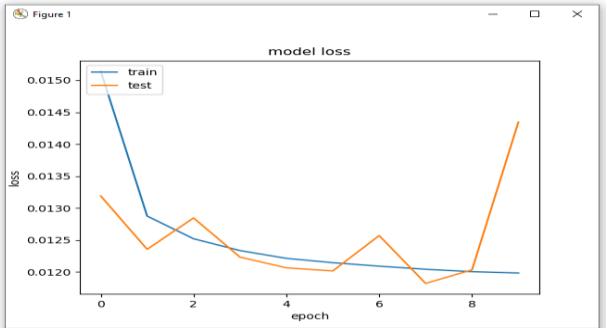

Fig. 5. Denoising Autoencoder Model Loss. Figure depicts the result of the loss, of every single- epoch sequence of Denoising Autoencoder model. It is found that the loss value is getting reduced based on the number of epoch sequence. The activation function used is Relu and loss function is MSE.

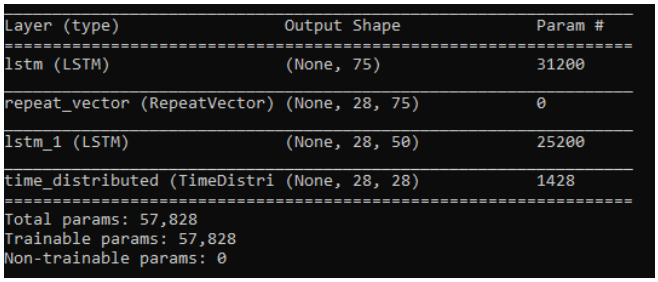

Fig. 6. Shows the total parameters used by the LSTM model.

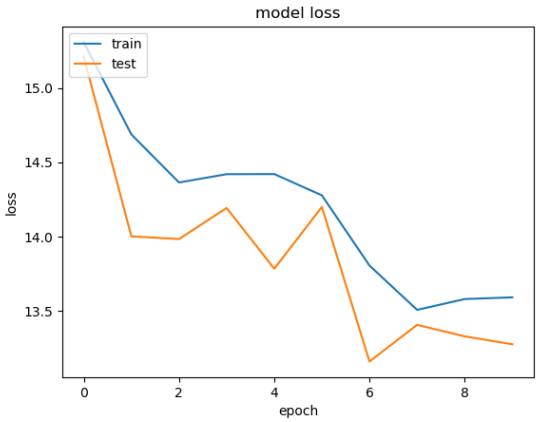

Fig. 7. LSTM Model Loss, where the loss function used is MSE and activation function is relu.

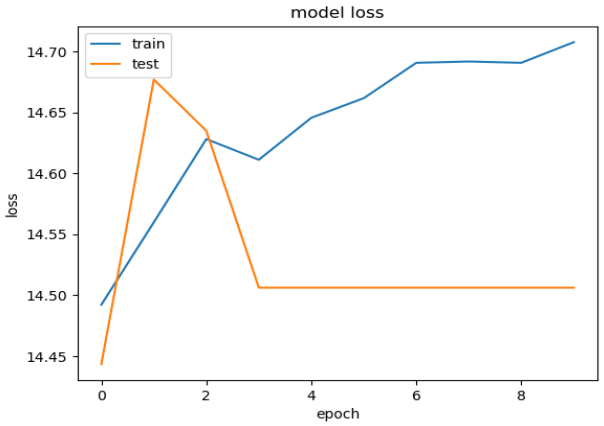

Fig. 8. is the model loss of LSTM combined with Denoising Autoencoders. Figure shows that based on the number of epochs, the loss value varies. The loss function used is MSE, activation function is relu.

V. CONCLUSION

In this paper three different models have been developed. From the experimental results, it was shown that the performance of LSTM+Denoising Autoencoder is based on the quality of the predicted frame and on the number of epochs. The loss value can be reduced by increasing the epochs. The given method improves the accuracy compared to the LSTM model and it is observed that the accuracy of the given method achieves up to 33% over LSTM model.

This work can be extended further with variations of LSTM Models and Autoencoders.

REFERENCES

1. Nitish Srivastav, Elman Mansimov, Ruslan Salakhutdinov, "Unsupervised Learning of Video Representations using LSTMs", ICML, arXiv:1502.04681v3 [cs.LG] 4 Jan 2016
2. M Mathieu, C Couprie, Y LeCun , "Deep multi-scale video prediction beyond mean square error", Published in ICLR 2015, arXiv preprint arXiv:1511.05440
3. William Lotter, Gabriel Kreiman & David Cox, "Unsupervised Learning Of Visual Structure Using Predictive Generative Networks", Published in ICLR 2016,arXiv:1511.06380v2 [cs.LG].
4. Mukherjee, Subham & Ghosh, Spandan & Ghosh, Souvik & Kumar, Pradeep & Roy, Partha. Predicting Video-frames Using Encoder-convlstm Combination. 2027-2031. 10.1109/ICASSP.2019.8682158, IEEE 2019.
5. Wei Henglai, Xiaochuan Yin and Penghong Lin, "Novel Video Prediction For Large-Scale Scene Using Optical Flow".arXiv: ABS/1805.12243 (2018).
6. Liang, Xiaodan, Lisa Lee, Wei Dai and Eric P. Xing, "Dual Motion GAN for Future-Flow Embedded Video Prediction," 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017.
7. W Byeon, Q Wang, RK Srivastava, P Koumoutsakos, "ContextVP: Fully Context-Aware Video Prediction",European Conference on Computer Vision
8. William Lotter and Gabriel Kreiman and David Cox," Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning" 2016, arXiv: 1605.08104[cs.LG]
9. Xu Jingwei, Ni Bingbing, Yang Xiaokang," Video Prediction via Selective Sampling", Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, Pages 1712--1722
10. Villegas, Ruben, Jimei Yang, Seunghoon Hong, Xunyu Lin and Honglak Lee. "Decomposing Motion and Content for Natural Video Sequence Prediction." (2017).
11. Wonmin Byeon & Qin Wang, Rupesh Kumar Srivastava, & Petros Koumoutsakos, "Fully Context-Aware Video Prediction" (2017)
12. Liu Ziwei, Raymond A. Yeh, Xiaoou Tang, Yiming Liu and Aseem Agarwala. "Video Frame Synthesis Using Deep Voxel Flow." *2017 IEEE International Conference on Computer Vision (ICCV)* (2017)
13. Marc Oliu, Javier Selva and Sergio Escalera. "Folded Recurrent Neural Networks for Future Video Prediction." *ECCV* (2018).
14. Villegas, Ruben, Jimei Yang, Yuliang Zou, Sungryull Sohn, Xunyu Lin and Honglak Lee. "Learning to Generate Long-term Future via Hierarchical Prediction." ICML (2017).
15. Pauline Luc, Camille Couprie, Yann Lecun, Jakob Verbeek. "Predicting Future Instance Segmentation by Forecasting Convolutional Features" ECCV- European Conference on Computer Vision, (2018).
16. Liang, Junwei & Jiang, Lu & Niebles, Juan Carlos & Hauptmann, Alexander & Li, Fei Fei. "Peeking into the Future: Predicting Future Person Activities and Locations in Videos" CVPR (2019).
17. Reda, Fitsum A., Guilin Liu, Kevin J. Shih, Robert Kirby, Jon Barker, David Tarjan, Andrew Tao and Bryan Catanzaro. "SDCNet: Video Prediction Using Spatially-Displaced Convolution." ECCV (2018).
18. Metz, Luke, Niru Maheswaranathan, Brian Cheung and Jascha Sohl-Dickstein. "Meta-Learning Update Rules for Unsupervised Representation Learning." *ICLR* (2019).
19. Emily L Denton et al., "Unsupervised learning of disentangled representations from video," in Advances in Neural Information Processing Systems, 2017.
20. Xianxu Hou, Linlin Shen, Ke Sun, and Guoping Qiu, "Deep feature consistent variational autoencoder," in Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on. IEEE, 2017.
21. Unsupervised Hierarchical Video Prediction conference paper at ICLR 2018.
22. Nelly Elsayed, Anthony S Maida, and Magdy Bayoumi. Empirical activation function effects on unsupervised convolutional lstm learning. In 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI), pages 336–343. IEEE, 2018
23. J. Hou, X. Wu, J. Chen, J. Luo, and Y. Jia, "Unsupervised deep learning ofmid level video representation for action recognition," in AAAI, 2018.
24. S. Lai, W.-S. Zhang, J.-F. Hu, and J. Zhang, "Global-local temporal saliency action prediction," IEEE Transactions on Image Processing, vol. 27, no. 5, pp. 2272–2285, 2018..
25. LSTM Networks-The math of Intelligent by Siraj Raval.
26. Y.-G. Jiang, Z. Wu, J. Wang, X. Xue, and S.-F. Chang, "Exploiting feature and class relationships in video categorization with regularized deep neural networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 2, pp. 352–364, 2018.
27. G. A. Sigurdsson, A. Gupta, C. Schmid, A. Farhadi, and K. Alahari, "Charades-ego: A large-scale dataset of paired third and first person videos," arXiv preprint arXiv: 1804.09626, 2018.
28. N.Lee,W.Choi,P.Vernaza,C.B.Choy,P.H.Torr,andM.Chandraker, "Desire: Distant future prediction in dynamic scenes with interacting agents," in CVPR, 2017.
29. D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in ICCV, 2015.
30. Y. Kong, Z. Tao, and Y. Fu, "Deep sequential context networks for action prediction," in CVPR, 2017.
31. https://towardsdatascience.com/denoising-autoencoders-explained-dbb82467fc2
32. P.Mohanaiah, p.Sathyanarayana, L.Gurukumar : Image Texture Feature Extraction using GLCM Approach, Volume 3, Issue 5,ISSN 2250-3153 (2013).
33. C. Feichtenhofer, A. Pinz, and R. P. Wildes, "Spatiotemporal multiplier networks for video action recognition," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017, pp. 7445–7454.

AUTHORS PROFILE

Mrs. Manju.D , Assistant Professor at GNITS, Hyderabad, has 14 yrs of teaching experience. Pursuing Ph.D from JNTUH, Completed M.Tech from HCU,Hyderabad.She has 8 papers in National and International conferences and in refereed journals.Her research interests include Image Processing, Artificial Intelligence and Data Mining.

Dr.Seetha.M, has teaching experience of 25 years which includes 8 years industrial experience. Madam received B.Tech from Nagarjuna University in 1992, M. S. from B I T S, Pilani in 1999 and Ph.D in Computer Science and Engineering in the area of image processing in December 2007 from Jawaharlal Nehru Technological University, Hyderabad. Currently she is working as Professor and Head of Department of CSE in GNITS, Hyderabad. Her research interests includes Image Processing, Neural Networks, Computer Networks and Data Mining. She is guiding 8 Ph.D scholars from JNTUH and JNTUK as Supervisor and 2 Ph.D scholars as co-supervisor from JNTUH. Two scholars are awarded with PH.D degree under her guidance. She has published 58 papers in referred journals and 72 papers in the proceedings of National/International Conferences and Symposiums.

Dr. Sammulal.P , Professor of CSE at JNTUH College of Engineering Jagtial (JNTUHCEJ).Sir has teaching experience of 17 years. He has obtained his BE in CSE from the Osmania University, Hyderabad, Telangana, India. M.Tech in CSE and PhD in CSE from the JNTUH University, Hyderabad, Telangana, India. His research interests includes Data Mining, Advanced Computing Techniques, Algorithms and Web Technologies. Sir has published many papers in reputed National and International Journals. Sir has published 2 books on Design and Analysis of Algorithms and Web Technologies and Applications. He is honoured with **"The Vishista Seva Puraskar" by J.N.T.U.H College of engineering Jagtial**, Nachupalli, for his meritorious services rendered to this Institution during 2009-10. He got Vittiya Saksharatha Abhiyan Best Campaign Award (VISAKA) in 2017 by MHRD, Govt. of India for his meritorious services as NSS Coordinator on Digital Transactions at the time of demonetization. He also honoured with Youth Worker Award in 2017 under NSS category.