

Analysis and Importance of Deep Learning for Video Aesthetic Assessments

Manish Bendale, Madhura V. Phatak, Dr. Nitin N. Pise

Department of Computer Engineering, MIT, Pune, Maharashtra, India

ABSTRACT

Deep Learning is one of the active analysis topic obtaining a great deal of analysis attention recently. This increase in analysis interest is driven by several area as that are being worked on like machine-based reality finding, good over-seeing, sensory activity recognition, online learning, world of advertisement, text analysis and so on. Videos have specific characteristics that make their method unique. Visual aesthetic typically: Remember what they see, understand and learn rather than what they hear. This paper principally emphasizes deep learning on basics of automatic video aesthetic assessments.

Keywords : Video Aesthetics, Preprocessing Techniques, Feature Extraction, Deep Learning, Applications.

I. INTRODUCTION

Computer technology are playing a very important and necessary role in our daily life. Mainly computer involved already in many tasks of gathering information using calculation, logic, we can say computers are very well predict complex events in the future like the real condition of weather, can able to manage complex systems and architecture like nuclear and atonic power plants, but still we do not take into account the state they are in the form of "living". Computers are very different from humans from the perspective of "thinking of mind". As for the person, feeling behind beauty could be a special kind of having thoughts; its Target is special outlines not forever in agreement like "Value". Because the underlying values become completely different, from our psychological thinking amendment. There's no great purpose or quality not like between different science like calculation or physics. So, to bridge over the limit of computers, the new research concept of "Deep Learning" is introduced by Walter Pitts and Warren McCulloch in 1943 [1].

It is to be noted that "Deep Learning" does not mingle with the standard of computers emotions. It would be imperative to first begin with human brain structures, to get a basic grasp on them and thus to outline the relation with computers. Then based on human psychology, the computers state of being are going to be put into use under the thought of Deep Learning and its completely different points of view.

II. DEEP LEARNING

2.1 How video aesthetics are related to Deep Neural Network?

Deep learning is a part of machine learning technology which has ability to learn and mimic the working principal of human brain. To mimic the brain working structure, deep neural network required large amount of datasets [2] to learn. Deep learning is a subset and the new research area of machine learning and its original goal is: Artificial Intelligence. Deep learning has ability to learn from the multiple levels of representation and the abstraction, these layers are helps in processing the amount of data like, images, videos, text, sound, etc.

Deep learning is a part of machine learning technology which has ability to learn and mimic the working principal of human brain. To mimic the brain working structure the deep learning technology required more number of datasets [3].

There is a different types of media aesthetics visions, video aesthetics is one of the aesthetics, due to the increasing in the popularization of high quality smart phones, the videos and image aesthetics are in great demands. In the current market the videos are playing important role in various fields like, advertisement, video conferencing, industrial representation etc. Previous works done using machine learning approaches, also called as handcrafted method in which, feature extraction were a common ways and methods for aesthetic analysis. CNN is recently used method to solve critical and complex problems as state-of-arts approaches [4].

2.2 Which are the features that can be identified?

Many other specific photographic rules can be used to discover particular aesthetic of videos by directly measuring physiological facts. This understanding and captured information is used to processed and extract meaningful information from the gathered data. In the earlier research, aesthetics was defined by Handcrafted and learned features [6].

The aesthetic quality of images and videos are judged by their low, middle and high level properties. The low level properties of an image are that of color, texture, edges and intensity [7]. The middle level property is the object in the image or video. And the high level properties are the photographic rules, mainly comprising of the Rules of Thirds (RoT), Visual Balance (VB), Diagonal Dominance (DD), Simplicity and the Depth of Field (DoF) [7].

Aesthetics assessment is a subjective field [4] because individual preferences differ according to personal taste too, thus what may be pleasing to one person may not be pleasing to the other. Therefore different communities define aesthetics differently, based on psychological and emotional aspects [8].



Fig. 1 High and Low Lighting Effect



Fig. 2 Low and High color contrast



Fig. 3. Depth of Field



Fig. 4 Rule of Third



Fig. 5 Sharp and Smooth Texture

2.3 Comparisons of Deep Learning with previous method about Video Aesthetic Assessment

All the present system is based on the handcrafted methodologies as we find in literature. Hand-crafted feature for images and videos aesthetics is good feature extractors based on user preferences algorithms, to design this features required a considerable amount of engineering skill and domain expertise. The main limitation is the excessive time consumption and the accuracies of aesthetic assessments. There are many researches done on the image aesthetics using hand crafted as well as deep learning methods [16].

Video aesthetics using deep learning, there exists a scope for future research as this field is less explored [19]. The table (table 1) shows the difference between the hand crafted and deep learning technology:

Handcrafted Method	Deep Learning Method
Handcrafted features judge by commonly established	Deep learning method work on highly and accurate

photographic rules, mainly low and middle level features.	feature extractions by using hidden layer concept.
This is sequential step-by-step method and time consuming to extract and classification of features.	This is parallel method using hidden layers concept and required less amount of time.
This method require engineering skills and human experience and do not need large dataset to train network.	This method depends on the amount of dataset to train the network for accuracy.
Backpropagation is not allowed to correct the final results.	Backpropagation is main advantage of CNN.
Classification done at once only.	Allow to use more fully connected layer for classification.

Table 1. Difference between the handcrafted and deep learning method.

There are the many researcher working on the deep learning to improve more accurate result in different field also. In the today life, the video aesthetic is the important multimedia application. Many field use the video for improver the selling quantity of the products. So, the video aesthetic is self-increasing area of aesthetic and in the future, it will go on the great demand. The deep learning provides a good opportunity for the video aesthetic assessment to improve the aesthetic time and the accuracy of video aesthetic editing. In this method, we deploy the multiple method of photography on the neural network and extract the good feature using the hidden layer of neural network.

2.4 Areas of Deep Learning related to video aesthetic assessments

- Detecting and recognizing aesthetic information
Aesthetics is a study that relates to the relationship between the mind and emotions in assessing beauty [8]. The aim of aesthetics analysis is to define the science in assessing the aspects of arts and beauty sought in images and videos. Human brain thinking structure begins with gathering information from different environment which take fact about users psychology and understanding with behavior. The fact gathered are of the form of features which are contains in a video and this features used to perceive people understanding behind videos beauty. For example, a camera might take pictures of the any object; Then human mind start its working and identify the object position, color contrast between foreground and background, Depth of Field, and etc.

- Psychology in machine

Another area within video aesthetic computing is that the design of procedure systems to exhibit either innate robust mind psychological powers or that is in a position of as definitely acting the part of aesthetics. An additional helpful move involving computers based mostly on current technology based powers, is that the simulation of aesthetic in taking pleasure in communication agents to boost interactivity between human like and machine like communications. Thinking ability in machines could be connected with define states connected with forward development (or exist without of forward development) of a pc primarily based learning system supported rules [8].

2.5 Another applications related to video aesthetic

Consider online-learning applications based on computers wherein computing the users feelings can be used to adjust the conference learning, content can be more polished and tasteful form may be used while teaching when a learner is uninterested, or paused. Human Psychological being healthy,

suggestions behind aesthetic, can help to the users, based on current emotion behind video beauty by computing applications when working out a particular user's state of feelings [9].

Visual aesthetic has possible applications in human computer interaction, now a day human computer interaction is varying area in different fields. Many researchers are working on gesture technology, in which camera capture the video of moving body gesture and according to movement of gesture the machine will understand, what action required on the basis of gesture. Virtual reality is also a greater area in human computer interaction, which shows the 3D video and create environment according to human psychology and emotion at the time of interactions.

Video aesthetics in Advertisements [10]: Now a day, the increasing in the population the requirement of the peoples goes on high demand. Due to renewed and changing consumer demands and the rapidly developing technological factors, companies need to search of new strategies to make a difference in their products. As competition increased marketers started to focus on new approaches for product to attract consumers' perception and attention. One of the most effective ways of differentiating is using video aesthetics.

In gaming area a mistake that's unfortunately only too common within the gaming business, is to lump video graphics and video aesthetics together as if they were the same factor. This, unnecessary to say, is false. The distinction becomes obvious once one compares, as an example, a heavily artificial, cartoon game that aims for realism specially else. The former's video aesthetic, despite its presumptively smaller budget and scope, can possibly hold up longer than the latter's.

Video aesthetic is also being applied to the growth of communicative technologies for use by people [10].

2.6 Different Algorithms Employed by Researchers

In the work on regression and classifying video aesthetic, the most frequently used classifier algorithms are Artificial Neural Networks (ANN), Deep Neural Network, Hidden Markov Models (HMMs), Gaussian Mixture Model (GMM), Support Vector Machines (SVM), k-nearest neighbor (k-NN), and Decision Tree Algorithms. Numerous studies have decisively showed that choosing the acceptable classifier will meaningfully improve the performance of the system.

Below gives a brief description about algorithm:

- K-NN – Classification prediction by defining the object in the feature space, and comparing this object with “k” nearest neighbors. In which the majority vote predict the classification of gathered data.
- GMM – it is a probabilistic based model used for signifying and classification of the presence of sub part within the whole pats. Every sub part is outlined exploitation the mixture distribution which permits for classification of explanations into the sub-parts [27].
- SVM – is used to binary linear classification and it is also used for regression, which selects in the two (or more) different classes, each input may fall into this class [28].
- ANN – is a mathematical model, which has ability to mimic brain structure of working process, it is enthused by biological neural networks that can work and extract feature similar to human mind [29].
- Decision Tree algorithms – are supported succeeding a decision tree throughout that the leaves of the tree represent the classification result, and branches represent the conjunction of consecutive features that lead to the classification [30].
- HMMs – An applied mathematics markov model during which the conditions and state transitions are unobtainable for the observation.

In its place, the series of outputs dependent on the states are visible [31].

2.7 Artificial Intelligence and Deep Learning

In 'Streaming Media magazine by Nadine Krefetz [33]', defined Video AI (artificial intelligence) has the capability to resolve variety of time-consuming, video-related issues with automation. However that doesn't mean its supernatural powers which will exclude human brain thinking behind beauty control. To supply a way of where video AI is in early 2018, what follows are variety of real-life examples during which AI helps to feature structure to the unstructured world of video aesthetic assessment.

At current years, there are many hype of applied science and especially within the field of artificial intelligence. Inspiring the machines to find out from the information which are gathered from different source and build accurate decisions within the areas like in the field of education, in medical and healthcare industries, finance and etc. There are multiple applications which are the mix of each artificial intelligence and Deep Learning.



Fig. 6 Surveillance with Deep learning and AI

Surveillance and security are the major concern of daily life, now a day, and every field contain video camera, with the intelligence in which when any unwanted event happened then camera automatically operate and collect the large amount of data continuously. Internally this complex data are

recognize and classify by the deep learning method to predict accurate results.

Other applications are like advertisement [20] and online learning, is great in demands, and video aesthetic has ability to make perfect communication between the industries and their client. And for the child, learning environment creation according to child psychology is great aspects of AL and deep learning in video assessment areas.

III. FUTURE SCOPE

3.1 Deep learning and Market of video aesthetic application

It is likely that the Deep Learning market will revenue in the Video Aesthetic Advertising segment amounts to US\$27,799m in 2018. Revenue is expected to show an annual growth rate (CAGR 2018-2023) of 13.7%, resulting in a market volume of US\$52,760m by 2023. The average revenue per Internet user currently amounts to US\$9.34 [34].

Deep Learning is about to revolutionize the approach organizations, particularly across sectors like retail, healthcare, government & defense, and academe, gather, organize, collaborate, and convey data. This is often a large rising within the world IT market and holds an awfully strong growth potential.

This currently innovative technology permits devices or machines to assemble, predict, resolve complexity, scrutinize, and decide future outcomes and be sort of a sophisticated and intelligent system, like human brain working structure. The system captures all dimensions of sensory vertex and different physiological and psychological changes in subjects/people beneath observation. Deep Learning mechanism relies on the component of the human brain that plays a large role in higher psychological or mental process by examining the behavior of a human's thinking and taking choices and decision on the concept of it slow series analysis [22].The Deep Learning systems incorporate technologies as

well as contact primarily based and non-contact technologies and captures human brain calculation through software system ways like neural analytics, gesture recognition, speech recognition, natural language processing and different physiological recognition features, to capture, archive, and discuss with existing knowledge [25].

3.2 Comparison between The Gartner Hype Cycle

"The Gartner hype Cycle for emerging Skills" as shown in figure 4 is that the broadest hype Cycle, which incorporates that includes technologies that are the stress of attention due to significantly nice levels of interest, and people that Gartner believes have the potential for important impression. Betsy Burton distinguished analyst at Gartner says, "We encourage CIOs and alternative IT front-runners to devote time and energy focused on innovation, instead of simply business progression, whereas additionally gaining inspiration by scanning on the far side the bounds of their trade."

"Deep Learning" isn't on the graph for the first time since 2015", on that time people do not motivated about the Deep learning technology(see figure 7). In the 2018 Garter hype Cycle for rising Technologies (see figure 8) the location of "Deep Learning" within the transition section is on the "Peak" whereby a lot of analysis focus is driving expectations from the sector on top of ever before and within the next decade it's expected to peak because the technology matures and analysis outcomes are reported from varied analysis teams [32].

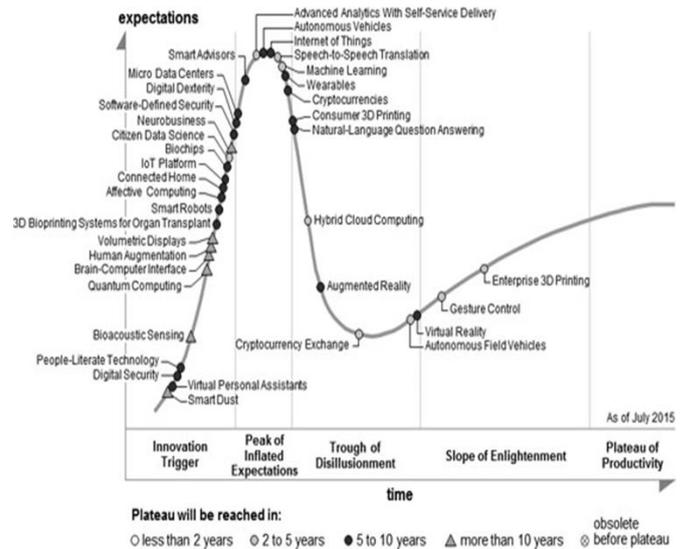


Fig 7. Gartner Hype Cycle 2015

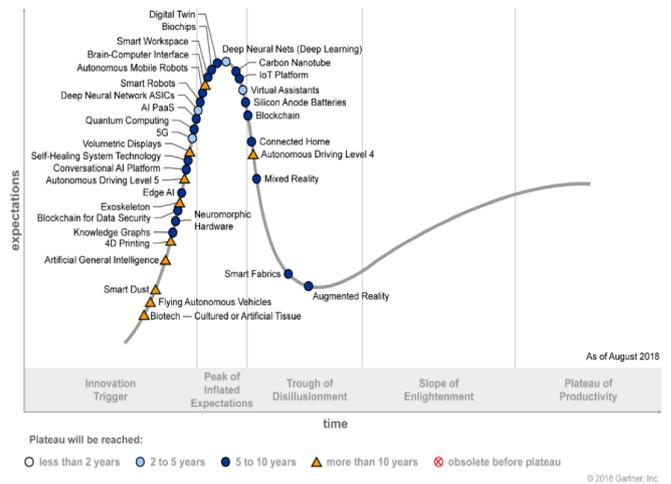


Fig 8. Gartner Hype Cycle 2018

IV. CONCLUSION

Aesthetic assessment of videos and the images is a challenging field. Many researchers are working on the classification techniques and they find that there is no foolproof technique for the classification and prediction [2]. The field of aesthetic assessment is subjective [17] and thus deep learning provides a multi-level learning method to get accurate classification [22]. Here deep learning provides the automatic aesthetic assessment for videos and images [10]. The image features provided for the video motion feature are extraction, which can provide to be broadly applicable for video aesthetic

assessments. The frame conversion and the shots detection is help to understand the characteristics of video motion to good shot quality of videos. In the future, the feature extraction and the classification both using the deep learning to take more accurate result on aesthetics. Many editor are working on the animation work in the videos, using deep learning we can modify and get the better result over that videos. The video transmission is based on the video size, so we can use the deep learning base aesthetic video transmission with less size and high aesthetic assessments.

V. REFERENCES

- [1]. Dhiraj Joshi, Ritendra Datta, Elena Fedorovskaya, Quang-Tuan Luong, James Z. Wang, Jia Li, and Jiebo Luo, "Aesthetics and Emotions in Images", IEEE Signal Processing Magazine, Volume: 28, Issue: 5, Sept. 2011
- [2]. Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, and James. Z. Wang, "Rating Image Aesthetics using Deep Learning ", IEEE Transactions on Multimedia, Volume: 17, Issue: 11, Nov. 2015
- [3]. Yubin Deng, Chen Change Loy, and Xiaoou Tang, "Image Aesthetic Assessment", IEEE Signal Processing Magazine, Digital Object Identifier 10.1109/MSP.2017.2696576, 11 July 2017.
- [4]. Yubin Deng, Chen Change Loy, and Xiaoou Tang, "Image Aesthetic Assessment: An Experimental Survey", arXiv:1610.00838v2 [cs.CV] 20 Apr 2017.
- [5]. Ritendra Datta Dhiraj Joshi Jia Li James Z. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach", Computer Science, vol 3953. Springer, Berlin, Heidelberg, ECCV 2006
- [6]. Chongke Bi ,Ye Yuan, Ronghui Zhang , Yiqing Xiang,Yuehuan Wang, and Jiawan Zhang, "A Dynamic Mode Decomposition Based Edge Detection Method for Art Images", IEEE Photonics Journal (Volume: 9, Issue: 6, Dec. 2017
- [7]. Roberto Gallea, Edoardo Ardizzone, and Roberto Pirrone, "Automatic Aesthetic Photo Composition", Image Analysis and Processing – ICIAP 2013: 17th International Conference, Naples, Italy, September 9, 2013.
- [8]. Stephen Groening, "Introduction: The Aesthetics of Online Videos", *The Aesthetics of Online Videos*, June 2016, DOI: <http://dx.doi.org/10.3998/fc.13761232.004.0.201>
- [9]. christos tzelepis, eftichia mavridaki, vasileios mezaris, ioannis patras, "Video aesthetic quality assessment using kernel support vector machine with isotropic gaussian sample uncertainty (ksvm-igsu)", IEEE Int. Conf. On Image Processing (Icip 2016), Phoenix, Az, Usa.
- [10]. Jun-Ho Choi, Jong-Seok Lee, "Automated Video Editing for Aesthetic Quality Improvement", 23rd ACM international conference on Multimedia, Brisbane, Australia — October 26 - 30, 2015
- [11]. Anselm Brachmann, Christoph Redies, "Computational and Experimental Approaches to Visual Aesthetics", Front. Comput. Neurosci., 14 November 2017 | <https://doi.org/10.3389/fncom.2017.00102>.
- [12]. Aarushi Gaur, "Ranking Images Based on Aesthetic Qualities", Submitted for the Degree of Doctor of Philosophy from the University of Surrey, April 2015.
- [13]. Loris Nanni, Stefano Ghidoni, Sheryl Brahma, "Handcrafted vs. non-handcrafted features for computer vision classification", Elsevier Journal on Pattern Recognition 71 (2017) 158–172, 2017.
- [14]. Hsin-Ho Yeh, Chun-Yu Yang, Ming-Sui Lee, and Chu-Song Chen, "Video Aesthetic Quality Assessment by Temporal Integration of Photo-

- and Motion-Based Features”, IEEE Transactions on Multimedia, Vol. 15, No. 8, December 2013.
- [15]. Katharina Schwarz, Patrick Wieschollek, Hendrik P. A. Lensch, “Will People Like Your Image? Learning the Aesthetic Space”, arXiv:1611.05203v2 [cs.CV] 4 Dec 2017
- [16]. Yanran Wang, Qi Dai, Rui Feng, Yu-Gang Jiang, “Beauty is Here: Evaluating Aesthetics in Videos Using Multimodal Features and Free Training Data”, ACM international conference on Multimedia, Pages 369-372, Barcelona, Spain — October 21 - 25, 2013
- [17]. Yuzhen Niu, Feng Liu, “What Makes a Professional Video? A Computational Aesthetics Approach”, IEEE Transactions on Circuits and Systems for Video Technology, Volume: 22, Issue: 7, July 2012
- [18]. Subhabrata Bhattacharya, Behnaz Nojavanasghari, Tao Chen, “Towards a Comprehensive Computational Model for Aesthetic Assessment of Videos”, ACM 978-1-4503-2404-5, 2013, Barcelona, Spain, <http://dx.doi.org/10.1145/2502081.2508119>.
- [19]. Chun-Yu Yang, Hsin-Ho Yeh, and Chu-Song Chen, “Video Aesthetic Quality Assessment By Combining Semantically Independent And Dependent Features”, Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference, 2011, Prague, Czech Republic.
- [20]. Yanxiang Chen, Yuxing Hu, Luming Zhang, Ping Li, Chao Zhang, “Engineering Deep Representations for Modeling Aesthetic Perception”, IEEE Transactions on Cybernetics, Volume: PP, Issue: 99, 11 December 2017.
- [21]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems (NIPS), pp. 1106–1114, 2012.
- [22]. D. Ciresan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3642–3649, 2012.
- [23]. Y. Sun, X. Wang, and X. Tang, “Hybrid deep learning for face verification,” in The IEEE International Conference on Computer Vision (ICCV), 2013.
- [24]. Jingwei Xu, Li Song, Rong Xie, “Shot boundary detection using convolutional neural networks”, *Visual Communications and Image Processing (VCIP) 2016*, pp. 1-4, 2016.
- [25]. Simone Bianco, Luigi Celona, Paolo Napoletano, and Raimondo Schettini, “Predicting Image Aesthetics with Deep Learning” ,Springer International Journal on Advanced Concepts for Intelligent Vision Systems, 21 October 2016, volume: 10016. Springer, Cham.
- [26]. C. D. Manning and H. Schütze, “Foundations of Natural Language Processing,” Reading, pp. 678, 2000.
- [27]. P. K. Abhishek Kumar Chauhan, “Moving Object Tracking using Gaussian Mixture Model and Optical Flow,” vol. 3, no. 4, pp. 243–246, 2013.
- [28]. D. K. Srivastava, L. Bhambhu, and B. Cet, “Data classification using support vector machine,” Journal of Theoretical and Applied Information Technology, 2010.
- [29]. M. S. and W. P, “Research Paper on Basic of Artificial Neural Network,” International Journal of Recent Innovation Trends Computer Communication, vol. 2, no. 1, pp. 96–100, 2014.
- [30]. P. Sharma and A. P. R. Bhartiya, “Implementation of Decision Tree Algorithm to Analysis the Performance,” Int. J. Adv. Res. Computer Communication. Eng., vol. 1, no. 10, pp. 861–864, 2012.
- [31]. F. Park, “The Hierarchical Hidden Markov Model : Analysis and Applications,” vol. 62, pp. 41–62, 1998.
- [32]. Gartner Hype Cycle:

- [33]. <https://medium.com/machine-learning-in-practice/deep-learnings-permanent-peak-on-gartner-s-hype-cycle-96157a1736e>
- [34]. NadineKrefetz , streaming media
<https://www.streamingmedia.com/Authors/6837-Nadine-Krefetz.htm>

Cite this article as :

Manish Bendale, Madhura V. Phatak, Dr. Nitin N. Pise, "Analysis and Importance of Deep Learning for Video Aesthetic Assessments", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 1, pp. 546-554, January-February 2019. Available at doi : <https://doi.org/10.32628/CSEIT1951100>
Journal URL : <http://ijsrcseit.com/CSEIT1951100>