Yanbu Journal of Engineering and Science

Volume 18 | Issue 1 Article 9

2021

Meme Generation Using Deep Neural Network to Engage Viewers on Social Media

Gasim Othamn Alandjani
Yanbu University College

Altaf Hussain Bouk
Department of Computer Science & Engineering, Yanbu University College

Follow this and additional works at: https://yjes.researchcommons.org/yjes

The journal in which this article appears is hosted on Digital Commons, an Elsevier platform.

Recommended Citation

Alandjani, Gasim Othamn and Bouk, Altaf Hussain (2021), Meme Generation Using Deep Neural Network to Engage Viewers on Social Media, *Yanbu Journal of Engineering and Science*: Vol. 18: Iss. 1, 81-87.
DOI: https://doi.org/10.53370/001c.29748
Available at: https://yjes.researchcommons.org/yjes/vol18/iss1/9

This Research Article is brought to you for free and open access by Yanbu Journal of Engineering and Science. It has been accepted for inclusion in Yanbu Journal of Engineering and Science by an authorized editor of Yanbu Journal of Engineering and Science.

Research article

Meme Generation Using Deep Neural Network to Engage Viewers on Social Media

Gasim Othamn Alandjani, Ph.D[1], Altaf Hussain Bouk, Ph.D.[1]

[1] Department of Computer Science & Engineering, Yanbu University College

Keywords: memes, generate memes, machine learning, deep feed-forward network

https://doi.org/10.53370/001c.29748

Yanbu Journal of Engineering and Science

Vol. 18, Issue 1, 2021

Abstract

In today's technological era, most new generations are using memes because the phenomenon of memes is rapidly gaining in popularity. In addition, various modifications and parodies use social and cultural phenomena to transform an original idea into another. Virus memes are sometimes called virus scripts and are the most widespread on the Internet. However, there is little scientific evidence for this assumption. This paper presents an innovative method to generate memes using machine learning algorithms that match the users' humorous and relevant captions. The VGG16 network is used to return the embedded image. We have collected feedback from citizens of more than five countries and received 37 feedbacks to find humorous memes. The obtained results from feedback are very encouraging from the generated memes.

INTRODUCTION

When users utilize various social networking websites, they have various questions, such as which image they should upload on Facebook that is more eye-catching. Which caption will be appropriate for specific alarming news that can be spread to as many users as possible? Which photographs should we use in a dog film so that we can reach a million views on YouTube? Daily, these questions bother professionals and regular internet users.[1-6] The ability to reach and be recognized by many people determines the influence of marketing techniques, social causes, ads, political campaigns, non-profit organizations, authors, and photographers, to name a few. The sample "Drake" meme utilized by many netizens is shown in Figure 1.

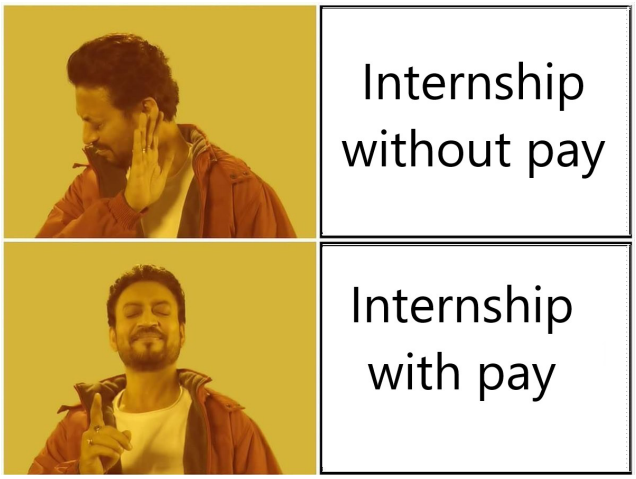

Figure 1. The sample of one of the popular copies of the 'Drake' meme on social media platforms.

Many factors can influence popularity when posting media to the Internet, including the time of day, the day of the week the user submitted the media, and the title used in the video. It is critical to determine whether pictures and media are transmitted through word of mouth.[1] Memes are one of the most popular forms of media among social media users. When users type a query into the Google search bar, Google Insight is likely to show overall data on the terms that have been input by users for that query. Google Insight gives weekly summaries based on how often a term/query has been searched/used since January 1st, 2004, and enables geographic and category filtering.

Overall statistics from throughout the world are recovered in this study. The data is normalized in such a way that the peak search activity for a query is scaled to a value of 100, although Google Insights never shows total search counts. As demonstrated in Figure 2 data acquired from Google Insights reflects relative search frequencies and does not allow for gauging absolute public interest in a topic. Because of its popularity among the younger generations, the graph of meme searches in Google trends is rapidly growing. A meme, according to social media networks, is an idea that spreads from one person to another within the same society.[2-7] These phenomena can sometimes extend globally, with a different topic and meaning in each culture.

Comprehending the mutations that are responsible for speeding or slowing the popularity of a meme can be influential in content creation and understanding the cultural makeup of online communities because many memes exist

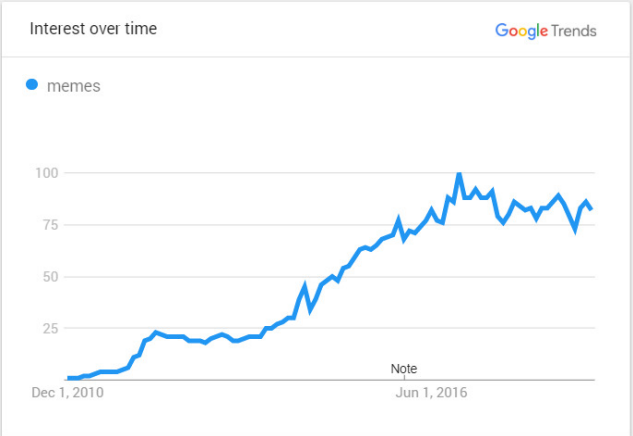

Figure 2. Google trends of the memes since 2010.

in the social network that persists by adjusting frequently.[8] The difficulty of distinguishing the underlying cultural meme from its different expressions in internet information is a problem with this area of research.[9]

Most memes are spread among internet users via social media sites, email, instant messaging systems such as WhatsApp, forums, and blogs. Images, news, video clips, and catchphrases are the most commonly shared types of material. Memes, to put it simply, are private jokes or snippets of current underground information that a large number of people are aware of. Memes evolve as a result of observation, imitation, or parody, or even a connected update in other broadcasting. Most Internet memes spread quickly, earning the term "viral script"; nevertheless, certain memes have been observed to go in and out of favor only a few times.

Memes are spread through a consensual, peer-to-peer mechanism rather than by force. Memes in the image-with-captions category are created utilizing a variety of machine learning algorithms. In this research, a new technique for creating memes with a witty caption is devised. No, the dataset is available for meme generation; therefore, we created a dataset with meme labels that can be used to create amusing memes for people of all cultures, genders, and ages. A CNN image embedding machine learning method was utilized in the first step of the picture caption system, followed by an LSTM RNN for meme production. We received 37 replies from over six different nations to determine the meme's quality and humor: South Korea, Pakistan, Nepal, Bangladesh, Saudi Arabia, and Vietnam. The suggested system is depicted in Figure 3 is a conceptual diagram. The picture caption model, the recurrent neural network for the language modeling challenge, and the RNN's attention mechanism in the associated task were all discussed in the following part. The model that was used. The approaches utilized to construct machine learning-based memes were detailed in the Approach section. The experiment and findings section that follows defines the actual user input that was evaluated, and the conclusion and future work section that follows makes conclusions and points in the right path.

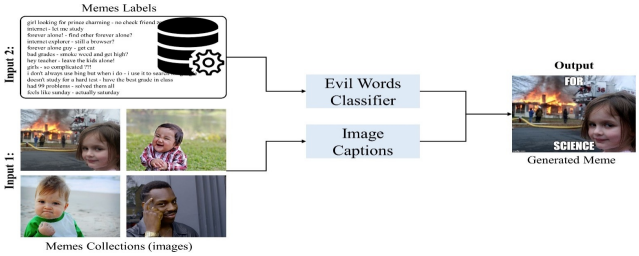

Figure 3. Conceptual diagram for the proposed memes method.

The rest of this paper is organized as follows. Section II briefly describes the background of the meme method and image captioning models. Section III explains the proposed method. Experimental results and discussion of the proposed method are described in Section IV. The paper is concluded in the last section.

RELATED WORK

Much research has attempted to categorize film genres from image posters using machine learning and deep learning. This section compares and contrasts studies that used movie posters to categorize genres and those that did not.

MEMES METHODS

Memes are gradually affecting social media and consuming the general public's feed. Researchers have only recently started looking into how visual memes and online users interact. Dupuis and his colleagues A survey was done to determine the personality traits of individuals who are prone to post picture memes with false information.[10] Crovitz and Moran[11] conducted a qualitative examination of visual memes used as a type of social media deception. Zannettou et al.[12] propose a large-scale quantitative measurement of picture meme spread on the Web, where the small, polarized community subreddit is racist and loathed on mainstream social media like Twitter and Reddit. We have found it to be particularly effective in pushing content.

Gal[13] observes that people use memes to express themselves and declare a position on a social topic, whether they support or oppose it. In the 2012 presidential election, Foster[14] looked into how memes were employed in a political setting to influence Facebook users. Work has been done on the automatic generation of memes,[15] but less on the analysis and classification of more detailed aspects such as the type and degree of built-in humor and aggression. Ratkiewicz et al.[16] want to utilize supervised learning based on characteristics and crowdsourced annotations collected from the topology of diffuse networks to detect memetic astroturfing on microblog streams. Ferrara et al.[17] By exploring pre-clustering ideas, we propose fully automated, unsupervised, and scalable real-time detection of memes in streaming data. Yoon[18] used critical discourse analysis in conjunction with multimodal discourse analysis to investigate racist online memes, but with the automatic and

computational classification of malevolent content inside memes. Detection still requires a lot of work.

IMAGE CAPTIONING MODELS

The process of generating a meaningful statement from an image is tough, yet it can have a significant impact, such as assisting visually challenged people in better understanding images. Picture captioning is a much more difficult task than image categorization, which has been the emphasis of the computer vision field. The relationship between the objects in a picture must be included in a description of the image. Aside from visual comprehension of the image, the above semantic knowledge must be communicated in a natural language such as English, which necessitates the use of a language model. Attempts to sew the two models together in the past have all failed. The creation of image captions is an important aspect of meme creation. Many different picture captioning approaches have been developed throughout the years. The architectures employed by the ILSVRC winners have made significant contributions to this discipline. One such architecture used by us was the VGG16 proposed by [19]. Apart from that, machine translation research has continually contributed to enhancing state-of-the-art performance in language technology.

Researchers at Microsoft's AI Lab used a pipeline approach to image captioning.[20] They created high-level characteristics for each putative object in the image using a CNN. They then utilized Multiple Instance Learning (MIL) to determine which region best corresponded to each word. On MSCOCO, the strategy received a 21.9% percent BLEU score. Following the pipeline technique, Google researchers developed the first end-to-end trainable model. The RNN paradigm employed in machine translation was a source of inspiration for them. Researcher[21] replaced this encoder RNN with CNN features of the image as the CNN features are widely used in all computer vision tasks. This model was dubbed Neural Image Caption by the researchers (NIC). Following this, two Stanford researchers tweaked the NIC. They learned about the inter-modal correspondences between language and visual data using a method that relies on image databases and sentence descriptions. Their multimodal embedding alignment strategy used a new combination of CNN over picture areas, bidirectional RNNs over words, and a defined objective to align the two modalities. They achieved state-of-the-art results using the Flickr8K, Flickr30K, and MSCOCO datasets.[22] Their model was further modified by [23] in 2015 when they proposed a dense captioning task in which each region of an image was detected and a set of descriptions generated. Another model which used a deep convolutional neural network (CNN) and two separate LSTM networks was proposed by [24] in the year 2016.

PROPOSED APPROACH

This section first explores the dataset reconstructed for the meme generation task and then explores the overall implementation of the proposed method.

DATASET

A dataset is an important thing for training any machine learning algorithm which is very important for the accuracy of the results. In this paper, two different datasets are used for the memes generation process. One is the Flicker8K dataset and the other is the dataset is generated by a python script from different websites. For the image captioning task, we have used the Flickr8k dataset which contains 8000 images with 5 captions per image. The dataset contains two folders one for images and one for text. Each image has a unique id and the caption for each of these images is stored corresponding to the respective id. Other datasets like Flickr30k and MSCOCO for image captioning exist but both these datasets have more than 30,000 images thus processing them becomes computationally very expensive. Captions generated using these datasets may prove to be better than the ones generated after training on Flickr8k because the dictionary of words used by the RNN decoder would be larger in the case of Flickr30k and MSCOCO.

For final memes generated using our dataset consists of approximately 2,500 images, labels, and caption triplets with 1200 unique image-label pairs, acquired from [25]. These memes are classified according to humor and ethics in the research community. Most of the memes on social networking sites have sexual and unethical words. Due to that those images and labels are classified and removed while generating the final meme. Labels are none but short descriptions mentioning to the image, i.e. the meme template, and are the same for identical images. Therefore, each image-label pair is allied with several around one hundred different captions.

PROPOSED NEURAL NETWORK

In the model proposed, we have used different algorithms that are combined into a single model which consists of a Convolutional Neural Network (CNN) encoder which helps in creating image encodings. We have to use VGG16 architecture with some modifications. These encoded images are then passed to an LSTM network which is a type of Recurrent Neural Network. The network architecture used for the LSTM network work similarly to the ones used in machine translators. The input to the network is an image that is first converted in a 224×224 dimension. We have used the Flickr8k dataset to train the model. The model outputs a generated caption based on the dictionary it forms from the tokens of the caption in the training set. The generated caption is compared with the human given caption via the BLEU score measure.

IMAGE CAPTIONING

In the suggested model, we merged several techniques into a single model that includes a CNN encoder that aids in the creation of picture encodings. We'll have to employ the VGG16 architecture with a few tweaks. After that, the encoded images are sent to an LSTM network, which is a form of Recurrent Neural Network. The LSTM network's net-

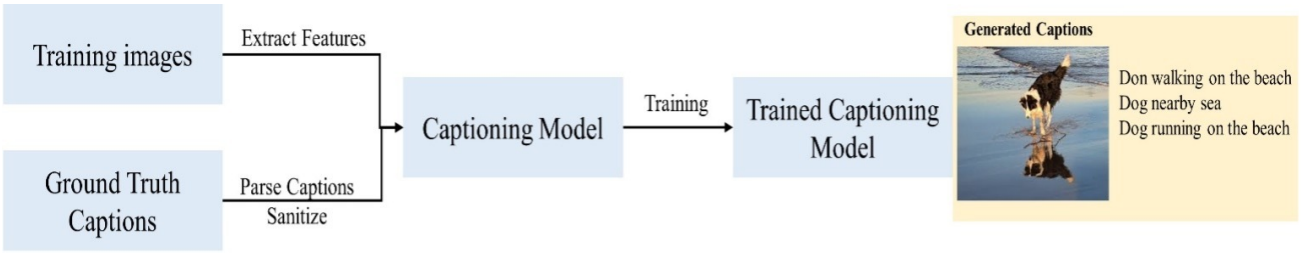

Figure 4. Conceptual diagram of image captioning.

work design is similar to that of machine translators. An image is fed into the network, which is then transformed to a 224×224 dimension. The model was trained using the Flickr8k dataset. Based on the dictionary it creates from the tokens of the caption in the training set, the model generates a caption. The produced caption is compared to the caption provided by a human using the BLEU score measure.

Consider the task of image classification as a separate problem at first. Various classifiers were used to classify the photos in the Cifar-10 dataset. Attempt to train the model using a K-Nearest Neighbor classifier first, followed by linear classifiers. Because a large loss factor at the moment of classification will amplify the loss, the accuracy of these models was significantly lower than expected. Even more so when it comes to caption creation. In addition, the model was trained using a basic Convolutional Neural Network and produced reasonable outcomes after only a few hours of training. And we discovered that CNN is a good strategy for using as an image encoder for the captioning model. Figure 4 depicts a conceptual schematic of the proposed image caption system architecture.

The multi-layer perceptron with the SoftMax activation function in the output layer is known as the completely connected layer. All of the neurons in the previous layer are connected to all of the neurons in the following layer as "fully connected". The convolution and pooling operations produce visual characteristics. The task of the fully connected layer is to map these feature vectors to the classes in the training data. The task of image classification on Cifar-10 has shown state-of-the-art results with the use of convents. Modified Alexnet is used and trained for images having 224×224 dimensions and hence need to be modified to be used for far-10 since the images in Cifar-10 are 32×32. We utilized a model with multiple layers of convolution and non-linearities.

The fully connected layer, which uses SoftMax activation to generate the scores of the 10 classes available in the Cifar-10 dataset, is used at the end of this study. The dataset on these convents has a training accuracy of 85% percent training. Figure 5 and Figure 6 show the plots of loss and accuracy on the test and validation sets. The model has been built using TensorFlow. Tensorflow is a machine learning open-source library created by the Google Brain team. Despite being a Python API, the majority of TensorFlow's code is developed in C++ and CUDA, Nvidia's GPU programming language. Because Python is slower than CPP, this aids TensorFlow in speedier code execution.

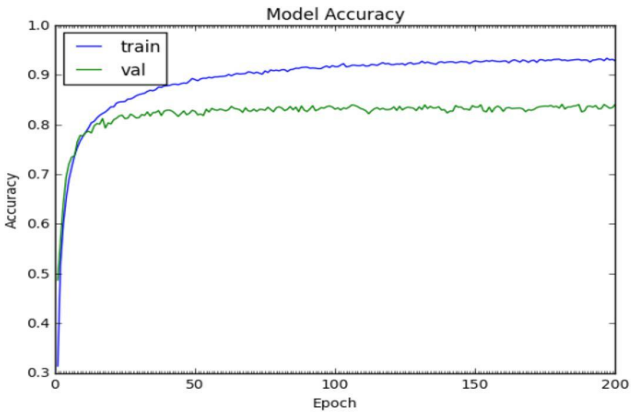

Figure 5. Accuracy and loss plot on the training and validation set.

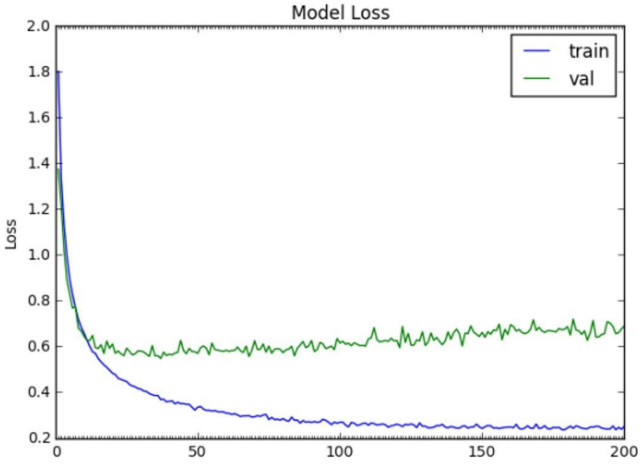

Figure 6. Accuracy and loss plot on training and validation set.

To calculate the BLUE score, create captions for all of the test photos first, and then utilize these captions as candidate sentences. We matched these candidate sentences to 5 human captions, averaging the BLEU score of candidates matching each of the references. Using Natural Language Toolkit (NLTK), a python library, we calculated 1000 BLEU scores for 1000 test photos.

Finally, we took the 1000 test photos and averaged the BLEU values. After 70 epochs of training with a batch size of 512, the model's net BLEU score was found to be 0.562, or 56.2 percent, while the state of the art on Flickr8k is about



Figure 7. Sample of final generated memes by our proposed method.

66 percent. By reducing the batch size, the net BLEU score can also be improved. In the majority of situations, the program was able to predict meaningful captions for an image.

EXPERIMENTS AND RESULTS

Due to the diversity of culture and human nature, determining the content a person enjoys is extremely difficult. As a result, numerous academics are working on content popularity prediction before releasing it on social media sites, blogs, websites, or other information sources. A feedback survey was designed in this study to find out consumers' actual ratings rather than expected evaluations. As a result, more than 37 responses from various religions, cultures, and genders have been gathered. More than six countries' social networking site users participate in this feedback and provide crucial comments on machine learning generated memes. The first seven memes were chosen for the feedback survey, which gathered user feedback. Figure 7 illustrates the first four memes.

After receiving valuable popularity feedback on the created memes, it was discovered that each meme received more than three out of five possible scores. As shown in Figure 8 and Figure 9. During the feedback, it was discovered that meme number seven was the most popular among female observers, and it is worth noting that none of the memes contain any aggressive or immoral language. And, most crucially, all memes are intended for a wide range of cultural and ethnic audiences, whether they like it or not. Blockchain technology can be used to extend the presented method.[25-27] Additionally, the proposed deep neural network can be used for other real-time applications such as vehicle classification.[28-31]

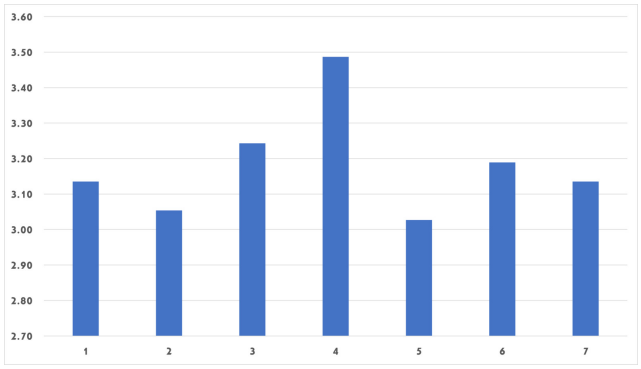

Figure 8. Results of the popularity feedback.

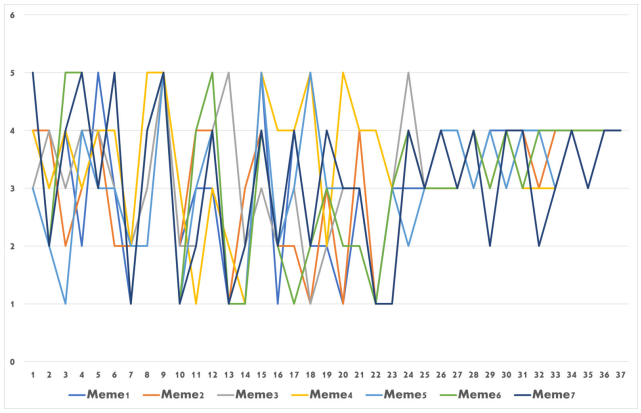

Figure 9. Results of the popularity feedback.

CONCLUSION

The creation of memes was proved in this paper using several networks. With this study, the model includes a convolutional neural network encoder and an LSTM decoder to aid in sentence generation. Experimenting with the model on the Flickr8K dataset yielded promising results. Based on the BLEU score, we assess the model's accuracy. The findings acquired from a variety of users utilizing a popularity feedback form show that memes may be created using machine learning and deep learning approaches. Furthermore, deep learning algorithms can be used to assess the popularity of memes before they are shared on social media networks. Machine learning could be used to generate GIF or video-based memes in the future. Furthermore, there are numerous memes on social media platforms that promote racism and hate speech; these memes may be eliminated from many internet platforms using machine learning algorithms, reducing hate speech, which is a major threat in many developing countries.

Submitted: September 30, 2021 AST, Accepted: October 30, 2021 AST

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at http://creativecommons.org/licenses/by/4.0 and legal code at http://creativecommons.org/licenses/by/4.0/legalcode for more information.

REFERENCES

1. Lakkaraju H, McAuley J, Leskovec J. What's in a name? understanding the interplay between titles, content, and communities in social media. In: *Seventh International AAAI Conference on Weblogs and Social Me.Ia*. ; 2013.

2. Mujtaba G. *Lightweight Client-Driven Personalized Multimedia Framework for next Generation Streaming Platforms [Ph.D. Dissertation]*. Ga- chon University, Seongnam, Korea; 2021.

3. Mujtaba G, Ryu ES. Human character-oriented animated gif generation framework. In: *2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC)*. ; 2021:1-6. doi:10.1109/majicc53071.2021.9526249

4. Mujtaba G, Ryu ES. Client-driven personalized trailer framework using thumbnail containers. *IEEE Access*. 2020;8:60 417-60 427. doi:10.1109/access.2020.2982992

5. Mujtaba G, Lee S, Kim J, Ryu ES. Client-driven animated gif generation framework using an acoustic feature. *Multimedia Tools and Applications*. Published online February 12, 2021. doi:10.1007/s11042-020-10236-6

6. Mujtaba G, Kim S, Park E, Kim S, Ryu J, Ryu ES. Client- driven animated keyframe generation system using music analysis. In: *Proceedings of the Korean Society of Broadcast Engineers Conference*. ; 2019:173-175.

7. Mujtaba G, Tahir M, Soomro MH. Energy efficient data encryption techniques in smartphones. *Wireless Personal Communications*. 2019;106(4):2023-2035,.

8. Coscia M. Average is Boring: How Similarity Kills a Meme's Success. *Scientific reports*. 2014;4(1):1-7,. doi:10.1038/srep06477

9. Shifman L. Memes in a digital world: Reconciling with a conceptual troublemaker. *Journal of computer-mediated communication*. 2013;18(3):362-377. doi:10.1111/jcc4.12013

10. Dupuis MJ, Williams A. The spread of disinformation on the web: An examination of memes on social networking. In: *2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE; 2019:1412-1418. doi:10.1109/smartworld-uic-atc-scalcom-iop-sci.2019.00256

11. Crovitz D, Moran C. Analyzing disruptive memes in an age of international interference. *English Journal*. 2020;109(4):62-69,.

12. Zannettou S, Caulfield T, Blackburn J, et al. On the origins of memes by means of fringe web communities. In: *Proceedings of the Internet Measurement Conference 2018*. ACM; 2018:188-202. doi:10.1145/3278532.3278550

13. Gal N, Shifman L, Kampf Z. "it gets better": Internet memes and the construction of collective identity. *New media & society*. 2016;18(8):1698-1714,.

14. Foster BJ. *It's All in a Meme: A Content Analysis of Memes Posted to 2012 Presidential Election Facebook Pages*. University of Arkansas; 2014.

15. Peirson V AL, Tolunay EM. Dank learning: Generating memes using deep neural networks [Preprint]. *arXiv*. Published online 2018.

16. Ratkiewicz J, Conover M, Meiss M, et al. Detecting and tracking the spread of astroturf memes in microblog streams [Preprint]. *arXiv*. Published online 2010.

17. Ferrara E, JafariAsbagh M, Varol O, Qazvinian V, Menczer F, Flammini A. Clustering memes in social media. In: *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASO.AM 2013)*. IEEE; 2013:548-555. doi:10.1145/2492517.2492530

18. Yoon I. Why is it not just a joke? analysis of internet memes associated with racism and hidden ideology of colorblindness. *Journal of Cultural Research in Art Education*. 2016;33.

19. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [Preprint]. *arXiv*. Published online 2014.

20. Fang H, Gupta S, Iandola F, et al. From captions to visual concepts and back. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2015:1473-1482. doi:10.1109/cvpr.2015.7298754

21. Vinyals O, Toshev A, Bengio S, Erhan D. Show and tell: A neural image caption generator. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2015:3156-3164. doi:10.1109/cvpr.2015.7298935

22. Karpathy A, Fei-Fei L. Deep visual-semantic alignments for generating image descriptions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE; 2015:3128-3137. doi:10.1109/cvpr.2015.7298932

23. Johnson J, Karpathy A, Fei-Fei L. Densecap: Fully convolutional localization networks for dense captioning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,*. IEEE; 2016:456-4574. doi:10.1109/cvpr.2016.494

24. Wang C, Yang H, Bartz C, Meinel C. Image captioning with deep bidirectional lstms. In: *Proceedings of the 24th ACM International Conference on Multimedia*. ACM; 2016:988-997. doi:10.1145/2964284.2964299

25. Alandjani G. Blockchain based auditable medical transaction scheme for organ transplant services. *3C Tecnología*. Published online November 6, 2019:41-63. doi:10.17993/3ctecno.2019.specialissue3.41-63

26. Mushtaq S, Gasim Dr, Farooq S, Nasser Dr, Adeel Dr, Shahbaz Dr. Hybrid Geo-Location Routing Protocol for Indoor and Outdoor Positioning Applications. *Int J Adv Comput Sci*. 2019;10(7):1-7. doi:10.14569/ijacsa.2019.0100701

27. Alandjani G. Features and potential security challenges for IoT enabled devices in smart city environment. *International Journal of Advanced Computer Science and Applications*. 2018;9(8):231-238,. doi:10.14569/ijacsa.2018.090830

28. Deng J, Cai J, Aftab MU, Khokhar MS, Kumar R, et al. Visual features with spatio-temporal-based fusion model for cross-dataset vehicle re-identification. *Electronics*. 2020;9(7):1083. doi:10.3390/electronics9071083

29. Cai J, Deng J, Khokhar MS, Aftab MU, et al. Vehicle classification based on deep convolutional neural networks model for traffic surveillance systems. In: *2018 15th International Computer Conference on Wav.Let Active Media Technology and Information Processing (IC- CWAMTIP)*. IEEE; 2018:224-227.

30. Jamali Z, Deng J, Cai J, Aftab MU, Hussain K. Minimizing vehicle re-identification dataset bias using effective data augmentation method. In: *2019 15th International Conference on Semantics, Knowledge, and.Grids (SKG)*. IEEE; 2019:127-130. doi:10.1109/skg49510.2019.00030

31. Zakria JD, Khokhar MS, Aftab MU, Cai J, Kumar R, Kumar J. Trends in vehicle re-identification past, present, and future: A comprehensive review. Published online 2021.