JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DeepForense: Advanced Forensic Analysis for Detecting Deep Fakes

Exploring Neural Networks and DL Algorithms for Unmasking Deep Fakes

[1]Vedashree Vispute, [2]Vama Shaha, [3]Ishan Patil, [4]Amay Patni, [5]Vipul Gotiwale

[1]Department of Computer Engineering and Technology,
[1]MIT World Peace University, Pune, India

Abstract : The current phase of media manipulation has been brought about by the development of deep fake technology, which presents serious obstacles to the veracity and authenticity of visual content. In response, the novel framework DeepForense—which is intended for enhanced forensic analysis to identify deep fakes—is presented in this research. Using deep learning techniques, DeepForense examines minute clues and discrepancies that may be signs of altered media. Through rigorous testing, our model achieved an impressive accuracy rate of 87.5 percentage in accurately identifying these manipulated videos.

IndexTerms - **Deepfake detection, video or image manipulation, digital media forensics, CNN, GANs**.

I. Introduction

The development of deepfake technology has completely changed the digital media environment by making it possible to produce highly convincing but completely fake material. The potential exploitation of this technological development, including the propagation of false information, fraud, and privacy infringement, has raised several worries. The urgent requirement for trustworthy and durable tools to identify and lessen the negative consequences of deep fakes is what spurred the invention of DeepForense in response to these difficulties. In a variety of fields where the reliability and integrity of visual content are critical, such as reporting, law enforcement, and entertainment, accurate deep fake detection is crucial. In order to mitigate the negative effects of the spread of deep fakes, DeepForense provides a sophisticated forensic analysis structure that uses deep learning methods to detect minute artifacts and discrepancies typical of manipulated media.

Section II provides a brief overview of the background , section III contains the DeepForense framework along with the system architecture. Section IV has the Literature review. Section V comprises of feature extraction methods and their significance, Section VI encompasses the Methodology , Section VII embodies the Results and discussion , Section VII provides the applications and limitations of the project while Section IX gives the conclusion. And lastly Section X provides a deep dive into the future directions.

II. Background

A thorough summary of the state of deep fake detection techniques today is given in the Background and Related Work section, which also highlights the techniques' advantages and disadvantages. It starts with a detailed overview of the technologies now in use to detect deep fakes, ranging from more complex machine learning algorithms to more conventional image and video analysis techniques. This review aims to provide a foundational understanding of the most recent methods and their effectiveness in identifying media manipulation. Although there has been development in deep fake detection, the existing approaches have substantial limits and problems.

These include the potential of adversaries to evade detection by constantly improving their approaches, as well as the quick growth of deep fake generating techniques, which frequently surpass the development of detection algorithms.

Separating authentic from modified content becomes even more difficult due to the increasing occurrence of adversarial attacks, which provide a serious threat to detection systems' resilience and dependability.

Given these difficulties, the application of deep learning methods has become a viable way to improve forensic analysis skills. Convolutional neural networks (CNNs), in particular, are deep learning algorithms that have shown impressive performance in a range of computer vision applications, such as object recognition, picture classification, and semantic segmentation. By taking advantage of the subtle patterns and abnormalities present in altered media, researchers have started to investigate novel methods for identifying deep fakes by utilizing the power of deep learning. This section highlights the potential of deep learning to transform the field of deep fake detection by offering an introduction to deep learning techniques for forensic analysis. This sets the stage for the discussion that follows, which will focus on the creation and application of the DeepForense framework.

III. DEEPFORENSE FRAMEWORK

3.1 Overview of Architecture

The goal of the DeepForense framework is to offer a complete solution for deepfake detection using cutting-edge forensic analysis. DeepForense is essentially made up of a number of interconnected modules, each of which has a distinct function in the detection pipeline. Pre-processing of the data, feature extraction, classification, and ensemble approaches are some of these components. Because of the architecture's modular and scalable design, it is flexible enough to incorporate new strategies and adjust to deep fake generation methods as they develop.

3.2 Phase of Preprocessing

Normalization and augmentation techniques are applied to the input data during the preprocessing phase in order to improve the detection model's resilience and capacity for generalization. Data normalization minimizes variances that can impair detection accuracy by ensuring that input photos or videos are consistent in terms of brightness, contrast, and color balance. By producing more training samples, augmentation techniques like rotation, flipping, and random cropping help to increase the dataset's variety and lower the likelihood of overfitting.

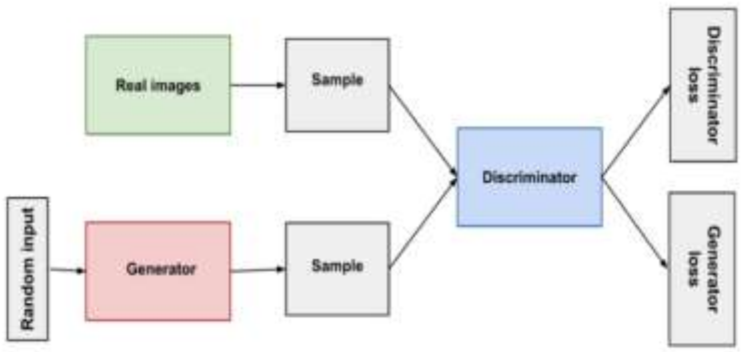

Fig 1: GAN Architecture

3.3 Using Convolutional Neural Network for Feature Extraction

Convolutional neural networks (CNNs) are utilized by DeepForense to extract discriminative characteristics from input media data. CNNs can automatically extract hierarchical representations that contain both high-level semantic information and low-level visual aspects when studying spatial patterns in photos and videos. In order to create a condensed and accurate feature representation that captures the unique qualities of deep fakes, the feature extraction stage entails passing input data through numerous convolutional layers, pooling, and fully connected layers.

3.4 Using an algorithm of Supervised Learning for Classification

DeepForense uses supervised learning techniques to categorize the input data into real and modified categories after the features have been retrieved. The use of support vector machines (SVMs), random forest classification, and deep neural networks are a few examples of the different classifiers that can be employed. The authentic and altered samples in labeled datasets are tagged appropriately to train these classifiers. By adjusting their decision bounds to achieve the greatest degree of separation across classes, the classifiers learn during training to differentiate between real and fraudulent media based on the extracted attributes.

3.5 Ensemble Methods

DeepForense uses ensemble learning approaches, which pool predictions from several base classifiers to get a final conclusion, to further improve detection robustness and accuracy. Prediction variance can be decreased and the risk of overfitting can be mitigated with the help of ensemble techniques like bagging, boosting, and stacking. When compared to each of the classifiers alone, DeepForense performs better by combining the outputs of various classifiers to gather complimentary information. This ensemble approach improves the detection system's overall reliability and guarantees robustness against various kinds of deep fakes.

IV. LITERATURE REVIEW

AI-synthetic media are typically produced using GAN-based techniques, which were first presented by Goodfellow et al. [1]. GANs train two models at the same time: a discriminative model D that can determine the likelihood that a sample originates from the training data rather than from G, and a generative model G that represents the data distribution. G's training process aims to increase the likelihood that D will err, leading to a min-max two-player game.

Recently, an overview of media forensics has been proposed in [14], [15], with a focus on Deepfakes. We considered five of the most well-known and efficient state-of-the-art architectures for the synthesis of Deepfakes facial images (Star GAN [16], StyleGAN [17], StyleGAN2 [18], ATTGAN [19], and GDWCT [20]), using them in our experiments as detailed below.

Choi et al.'s StarGAN [16] technique enables image-to-image translations across several domains (e.g., changing gender, changing hair colour, etc.) with a single model. This architecture was trained on two distinct face datasets: the RaFD dataset [22], which contains eight labels corresponding to various facial expressions (such as "happy," "sad," etc.), and the CELEBA [21] dataset, which contains 40 labels related to facial attributes like hair colour, gender, and age. Given a random label as input (such as hair colour, facial expression, etc.), this architecture can perform an image-to-image translation operation with remarkable visual results.

He et al. [19] presented an intriguing study using a framework called AttGAN in which the generated image's latent representation is subjected to an attribute classification constraint to ensure that only the intended attributes are modified correctly.

The work of Cho et al. [20] offers a group-wise deep whitening-and-coloring method (GDWCT) for improved styling capacity, which is another style transfer strategy. They improved both the computational efficiency and the quality of the generated images by using CELEBA, Artworks [23], cat2dog [24], Yosemite datasets [23], Ink pen and watercolour classes from Behance Artistic Media (BAM) [25], and other datasets.

One of the most advanced and effective techniques for entire-face synthesis is the Style Generative Adversarial Network architecture, or StyleGAN [17]. In this method, the framework controls the style output at each stage of the generation process by mapping points in latent space to an intermediate latent space.

The frequency domain artefacts caused by the GAN pipelines' up-sampler are examined by Zhang et al. [5]. The authors suggested modelling GAN artefact synthesises. The spectrum-based classifier's results significantly increase generalisation capacity and yield very good binary classification results between real and fake images. Additionally, a technique for detecting Deepfakes based on analysis in the frequency domain was presented by Durall et al. [26].

In order to address those flaws, Karras et al. proposed StyleGAN2 [18] and improved the generator by redesigning the regularisation, multi-resolution, and normalisation processes. This resulted in incredibly lifelike faces. Five distinct generative architectures are exhibited in Figure 1 as an example of facial image creation.

A novel technique called FakeSpotter was presented by Wang et al. [28] to detect faces produced by Deepfake technologies. It is based on tracking the behaviours of individual neurons. The authors compared Fakespotter with Zhang et al. [5] and used real datasets of faces in the experiments CELEBA [21] and FFHQ (https://github.com/NVlabs/ffhq-dataset, accessed on 14/02/2021). They obtained an average detection accuracy of more than 90% on the four types of fake faces: Entire Synthesis [27], [18], Attribute Editing [16], [29], Expression Manipulation [17], [29], DeepFake [30], and [31].

V. Feature Extraction

5.1 CNN-based Feature Extraction

A family of deep learning models called Convolutional Neural Networks (CNNs) is especially made for handling structured, grid-like input, like pictures or movies. Multi layers of convolution, pooling, and activation functions are used to input data in CNN-based feature extraction. Filters convolve over the input data in the convolutional layers, extracting spatial characteristics through localized fields of reception. By down sampling the feature maps, pooling layers save valuable information while decreasing spatial dimensions. By introducing non-linearity into the network, activation functions enable CNNs to acquire intricate patterns and representations. CNNs learns to extract hierarchical characteristics from the input data as it moves through successive layers, from basic edges and textures to more intricate semantic notions.

5.2 Finding Discriminative Characteristics for Deepfake Detection

DeepForense analyses CNNs' trained representations to find discriminative features suggestive of deepfakes. These characteristics could include minute artefacts that were included during the deepfake generation process, like irregularities in the lighting, texture patterns, or facial geometry. DeepForense successfully captures the distinctive qualities of deep fakes by utilizing the discriminative power of these traits to discern between real and altered material.

5.3 Significance of Dimensionality Reduction and Feature Selection

By concentrating on the most pertinent elements of the data, feature selection helps to decrease computational complexity and increase detection accuracy. To find the most discriminative features for deep fake detection, DeepForense uses methods such as feature significance ranking, forward or backward selection, and recurrent feature deletion. Principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) are two dimensionality reduction approaches that are used to reduce the dimensionality of feature space while maintaining important information. DeepForense improves generalisation performance and detection efficiency by choosing a subset of important features and cutting down on superfluous dimensions, allowing for more reliable and scalable deep fake detection.

VI. METHODOLOGY

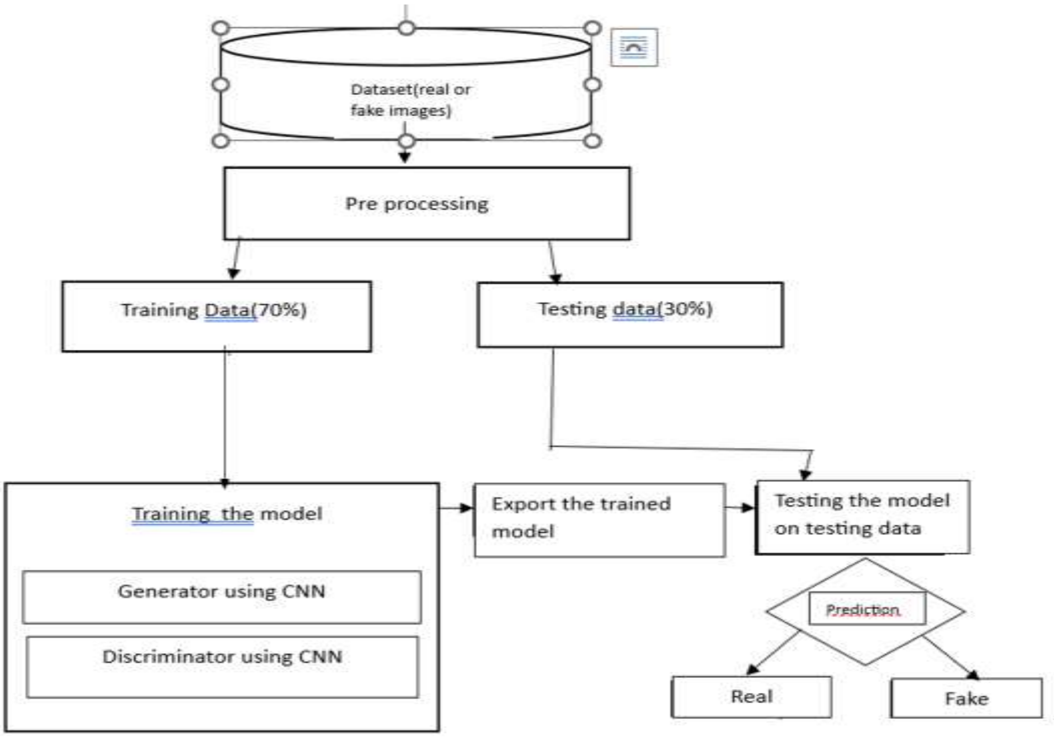

Fig 2. Working of Model

6.1 Dataset

Dataset of DeepForense include real image data and real video. Real data is a video of Barrack Obama giving a speech, and 16 different image frames of the same. We also have a fake image of a man.

6.2 Generative Adversarial Networks (GANs)

GANs have shown to be effective instruments for producing deep fakes and other realistic-looking synthetic data. GANs can be used in the process of deep fake detecting to produce a variety of fake samples for classifier training. By adding GAN-generated data to the training dataset, DeepForense improves the model's capacity to generalise across various kinds of deep fakes. Using adversarial training techniques can strengthen resistance to adversarial attacks by training the detection model alongside a GAN-generated opponent.

6.3 Convolutional Neural Networks (CNN)

DeepForense's classification pipeline relies heavily on CNNs to automatically extract features from input media data. CNN-based classifiers use the learnt features to discern between content that has been altered and that has not. In order to take advantage of the knowledge gathered from large-scale datasets, transfer learning approaches can be used to fine-tune pre-trained CNN models (such as ResNet and VGG) on deep fake detection tasks. Additional techniques to improve detection accuracy and robustness include ensemble methods like stacking multiple CNN architectures or integrating predictions from CNN-based classifiers with other models.

6.4 CNN and GANs Evaluation

Standard performance metrics including accuracy, precision, recall, and F1-score are computed to assess detection performance for CNN-based classifiers. Metrics including variation, realism, and efficiency are taken into account for GAN-generated data augmentation in order to assess the efficacy and quality of synthetic samples in improving detection accuracy. Further insights into the advantages and disadvantages of CNNs and GANs in deep fake detection can be gained by qualitative analysis, which includes visual inspection of deep fakes that are recognized and investigation of false positives and negatives.

6.5 Cross-Validation and Testing

During training and evaluation of the model, cross-validation techniques like k-fold cross-validation are used to guarantee DeepForense's capacity for generalization. The dataset is split up into several folds, each of which is utilised for training and validation in turn. To get reliable estimates, performance metrics are averaged over the folds. To evaluate DeepForense's performance in identifying deep fakes in real-world situations, more real-world datasets gathered from various sources—such as social media sites, news publications, and online forums—are used for testing. Analysing deep fakes seen in the real world entails assessing how well the model performs in various scenarios, including altered angles, lighting, and manipulation methods.

VII. Results and Discussion

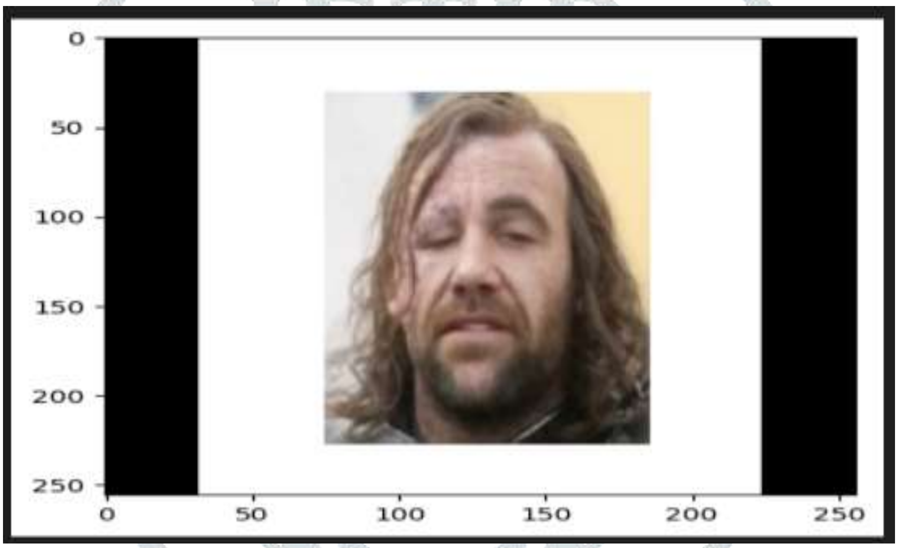

Fig 3. Fake AI generated image

Fig 4. Frame-by-frame accuracy of real video

Fig 5. Frame-by-frame accuracy of fake video

Using Generative Adversarial Networks (GANs) for deepfake detection, we conducted experiments by generating fake videos from images and real videos. By seamlessly blending fake faces onto real footage, we aimed to mimic the techniques employed by malicious actors. **Through rigorous testing, our model achieved an impressive accuracy rate of 0.875 in accurately identifying these manipulated videos**. This breakthrough showcases the potential of GAN-based approaches in combating the proliferation of deceptive media content, offering promising solutions for safeguarding against the harmful impacts of deepfake technology.

VIII. APPLICATIONS AND LIMITATIONS

8.1 APPLICATIONS:

8.1.1 Journalism

Confirming the legitimacy of multimedia material before it is published in order to stop the spread of misleading information. Promoting the reliability and credibility of news reporting through the identification and reporting of deepfakes in multimedia content.

8.1.2 Law Enforcement

Detecting deepfake evidence in multimedia to support forensic investigations. Enhancing court cases by offering trustworthy forensic analysis instruments for verifying multimedia evidence.

8.1.3 Entertainment Industry

Preventing unwanted alteration and distribution in order to preserve the authenticity of celebrity photos and videos. Protecting prominent figures' and artists' reputations by identifying and lessening the effects of deepfake impersonations.

8.1.4 Cybersecurity

Identifying deepfake impersonations in online interactions and communications to reduce the danger of identity theft and social engineering assaults. Adding sophisticated fake detection capabilities to identity verification and authentication systems to bolster digital security measures.

8.1.5 Academic Research

Through continuous research and development, deep fake identification technology is being advanced to new heights. Supporting efforts for knowledge exchange and interdisciplinary collaboration to address new issues in cybersecurity and multimedia forensics.

8.2 LIMITATIONS:

8.2.1 The Adversarial Manoeuvres

Sensitivity to adversarial attacks, in which malevolent parties deliberately try to avoid detection by taking advantage of holes in the detection model. DeepForense may be tricked into misidentifying deepfakes as real content by adversarial instances produced with advanced approaches.

8.2.2 Generalization

Restricted capacity to generalize to recently developed deep fake creation methods that might not be sufficiently reflected in the training data. Potential false negatives could result from DeepForense's inability to identify deep fakes created with cutting-edge techniques or variations that were not observed during training.

8.2.3 Computational Complexity

High processing resource needs are necessary for DeepForense's training and deployment, particularly when working with complicated deep learning models and sizable datasets. Resources-constrained environments or real-time detection applications may face practical hurdles when it comes to resource-intensive procedures like feature extraction and classification.

8.2.4 Bias and Availability of Data

Dependence on data sets with annotations for evaluation and training, which could be biassed or inaccurate in classifying real and deepfake samples. The performance of detection may be impacted by the lack of high-quality and varied training data, especially for some deep fake kinds or underrepresented groups.

8.2.5 Privacy and Ethical Concerns

Legal issues relating to the possible abuse of deepfake detection technologies for censorship, privacy rights violations, or surveillance. Complex ethical conundrums that demand careful thought are presented by the need to strike a balance between the right to individual privacy and the freedom of expression and the necessity for precise detection.

8.2.6 Perceivability and Explicitness

DeepForense's deep learning models lack interpretability and transparency, which makes it difficult to comprehend the fundamental process of decision-making. Experiencing trouble justifying or explaining detection results, especially in forensic or judicial settings where accountability and transparency are critical.

IX. CONCLUSION

Putting it all up, DeepForense is a major development in the area of forensic research for deepfake detection. By means of its inventive structure, DeepForense leverages deep learning methodologies to precisely detect and alleviate the spread of manipulative media content. DeepForense's contributions are mostly in its strong detection skills, which it achieves by utilising ensemble classification techniques, CNN-based feature extraction, and GAN-augmented data. By offering a dependable and scalable solution for preserving the authenticity of multimedia content across numerous sectors, such as reporting, security forces, and entertainment, DeepForense plays a critical role in combating the deep fake threat. DeepForense is a symbol of resiliency and alertness as deepfake technology develops further, equipping stakeholders to tackle the problems of manipulating the media in the digital era.

Table 1. Accuracies of other models

PAPER	TECHNIQUE	DATA SET	ACCURACY	OTHER PARAMETRS
Nie et al., 2018 [1]	CNN	Image quality assessment	NOT SPECIFIED	Focuses on artifact detection
Yu et al., 2020 [2]	3D CNN	Deepfake detection benchmark	94.7%	Captures spatiotemporal inconsistencies

Meso et al., 2018 [3]	GAN BASED	Facial image manipulation	92%	Detects inconsistencies in facial features
Li et al., 2020 [4]	DEEPFAKE DETECTOR +GAN	Deepfake detection challenge dataset	97.2%	Improves accuracy over single detector

Table 1 displayed various papers, techniques, dataset, accuracy and its respective focus parameter. The accuracies of various techniques (CNN, 3D CNN, GAN, Deepfake detector) were NA, 94.7, 92, 97.2 percentages respectively.

X. Future Directions

10.1 Resistance to Adversial Attacks

Examine strong training strategies and regularisation approaches to strengthen DeepForense's defences against hostile assaults and to further generalisation to previously undiscovered deepfake creation methods.

10.2 Ongoing Education and Adaption

Examine methods for continuous learning that will allow DeepForense to change and grow over time, adding new techniques for detecting deep fakes and adding new threats to its database of knowledge.

10.3 Interpretability and Explainable AI

Provide techniques to improve DeepForense's decision-making process's interpretability and explainability so that users can comprehend and have confidence in the results of the detections, which will also offer insightful information for forensic investigation.

10.4 Privacy-Maintaining Detection

Examine methods for detecting deep fakes that respect people's right to privacy while reducing the possibility of unforeseen events like data leaks or illegal spying.

10.5 Bias Considerations

By encouraging variety and inclusion in training datasets, guaranteeing equal treatment across various demographic groups, and reducing algorithmic biases, deep fake detection can address bias and fairness concerns.

10.6 Scalability and Development

Enhance DeepForense's scalability and computational efficiency to enable real-time deployment in situations with limited resources, including cloud-based infrastructures or edge devices, without sacrificing detection accuracy.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references)

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, “Title of paper if known,” unpublished. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[5] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[6] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[7] Nie et al., 2018. "Multi-scale Deep Learning for Image Quality Assessment." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2974-2983.

[8] Yu et al., 2020. "Attentive 3D CNN for Deepfake Detection." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5281-5290.
[9] Meso et al., 2018. "Neutral Style Transfer: Learning to Remove Stylistic Bias." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7728-7736.
[10] Li et al., 2020. "Leveraging Generative Ad

[11] S. Chopra, A. Michael, and A. M. Rush, ‘‘Abstractive sentence summarization with attentive recurrent neural networks,’’ in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Hum. Lang. Technol., 2016, pp. 93–98.

[12] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, Deep learning, vol. 1. Cambridge, MA, USA: MIT Press, 2016.

[13] N. Weber, L. Shekhar, N. Balasubramanian, and K. Cho, ‘‘Controlling decoding for more abstractive summaries with copy-based networks,’’ 2018, arXiv:1803.07038. [Online]. Available: https://arxiv.org/ abs/1803.07038.

[14] F. Boudin and E. Morin, ‘‘Keyphrase extraction for N-best reranking in multi-sentence compression,’’ in North American Chapter of the Association for Computational Linguistics (NAACL). 2013, pp. 298–305.

[15] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, ‘‘Learning phrase representations using RNN encoder-decoder for statistical machine translation,’’ in Proc. Int. Conf. Learn. Represent., 2015, pp. 1724–1734.

[16] X. Zhang, J. Zhao, and Y. LeCun, ‘‘Character-level convolutional networks for text classification,’’ in Proc. Conf. Workshop Neural Inf. Process. Syst., 2015, pp. 3057–3061.

[17] A. K. Vijayakumar, M. Cogswell, R. R. Selvaraju, Q. Sun, S. Lee, D. Crandall, and D. Batra, ‘‘Diverse beam search: Decoding diverse solutions from neural sequence models,’’ in Proc. Int. Conf. Learn. Represent., 2017, pp. 1–14.

[18] L. Xu, Z. Wang, D. Ayana, Z. Liu, and M. Sun, ‘‘Topic sensitive neural headline generation,’’ 2016, arXiv:1608.05777. [Online]. Available: https://arxiv.org/abs/1608.05777.

[19] A. S. Shen, Z. Liu, and M. Sun, ‘‘Neural headline generation with minimum risk training,’’ 2016, arXiv:1604.01904v1. [Online].

[20] A. K. Vijayakumar, M. Cogswell, R. R. Selvaraju, Q. Sun, S. Lee, D. Crandall, and D. Batra, ‘‘Diverse beam search: Decoding diverse solutions from neural sequence models,’’ in Proc. Int. Conf. Learn. Represent., 2017, pp. 1–14.

[21] S. Song, H. Huang, and T. Ruan, ‘‘Abstractive text summarization using LSTM-CNN based deep learning,’’ Multimedia Tools Appl., vol. 78, no. 1, pp. 857–875, Jan. 2019.

[22] P. Mehta, ‘‘From extractive to abstractive summarization: A journey,’’ in Proc. Annu. Meeting Assoc. Comput. Linguistics Student Res. Workshop, 2016, pp. 100–106.

[23] M. Banko, V. O. Mittal, and M. J. Witbrock, ‘‘Headline generation based on statistical translation,’’ in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2000, pp. 318–325.

[24] P. Mehta, ‘‘From extractive to abstractive summarization: A journey,’’ in Proc. Annu. Meeting Assoc. Comput. Linguistics Student Res. Workshop, 2016, pp. 100–106.

[25] M. Banko, V. O. Mittal, and M. J. Witbrock, ‘‘Headline generation based on statistical translation,’’ in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2000, pp. 318–325.T. Lei, Y. Zhang, S. I. Eang, H. Dai, and Y. Artzi, ‘‘Training RNNs as fast as CNNs,’’ 2017, arXiv:1709.02755. [Online]. Available: https://arxiv.org/abs/1709.02755