**INTRODUCTION**

The notable advances in artificial neural network (ANN) based technologies play an essential role in tampering with multimedia content. For example, AI-enabled software tools like Face App [1], and Fake App [2] have been used for realistic-looking face swapping in images and videos. This swapping mechanism allows anyone to alter the front look, hairstyle, gender, age, and other personal attributes. The propagation of these fake videos causes many anxieties and has become famous under the hood, Deep fake.

The term ``Deep fake'' is derived from ``Deep Learning (DL)'' and ``Fake,'' and it describes specific photo-realistic video or image contents created with DL's support. This word was named after an anonymous Redd it user in late 2017, who applied deep learning methods for replacing a person's .face in pornographic videos using another person's face and created photo-realistic fake videos. To generate such counterfeit videos, two neural networks: (i) a generative network and (ii) a discriminative network with a Face Swap technique were used [3], [4]. The generative network creates fake images using an encoder and a decoder. The discriminative network defines the authenticity of the newly generated images. The combination of these two networks is called Generative Adversarial Networks (GANs), proposed by Ian Good fellow [5].

Based on a yearly report [6] in Deep fake, DL researchers made several related breakthroughs in generative modeling. For example, computer vision researchers proposed a method known as Face2Face [7] for facial re-enactment.

This method transfers facial expressions from one person to a real digital 'avatar' in real-time. In 2017, researchers from UC Berkeley presented Cycle GAN [8] to transform images and videos into different styles. Another group of scholars from the University of Washington proposed a method to synchronize the lip movement in video with a speech from another source [9]. Finally, in November 2017, the term ``Deep fake'' emerged for sharing porn videos, in which celebrities' faces were swapped with the original ones. In January 2018, a Deepfake creation service was launched by various websites based on some private sponsors. After a month, several websites, including Gfycat [10], Pornhub, and Twitter, banned these services. However, considering the threats and potential risks in privacy vulnerabilities, the study of Deep fake emerged super fast. Rossler *et al.* introduced a vast video dataset to train the media forensic and Deep fake detection tools called Face Forensic [11] in March 2018. After a month, researchers at Stanford University published a method, ``Deep video portraits'' [12] that enables photo-realistic re-animation of portrait videos. UC Berkeley researchers developed another approach [13] for transferring a person's body movements to another person in the video. NVIDIA introduced a style-based generator architecture for GANs [14] for synthetic image generation. According to [6] report, Google search engine could find multiple web pages that contain Deep fake related videos (see Figure 1).We found the following additional information from this report [6]:

\_ The top 10 pornographic platforms posted 1,790C Deep fake videos, without concerning pornhub.com, which has removed 'Deep fakes' searches.

\_ Adult pages post 6,174 Deep fake videos with fake video content.

\_ 3 New platforms were devoted to distributing Deep fake pornography.

\_ In 2018, 902 articles were published in arXiv, including the keyword GAN either in titles or abstracts.

\_ 25 Papers published on this subject, including non-peer reviews, are investigated, and DARPA funded 12 of them.

Apart from Deep fake pornography, there are many other malicious or illegal uses of Deep fake, such as spreading misinformation, creating political instability, or various cybercrimes. To address such threats, the field of Deep fake

detection has attracted considerable attention from academics and experts during the last few years, resulting in many Deep fake detection techniques. There are also some efforts on surveying selected literature focusing on either detection methods or performance analysis. However, a more comprehensive overview of this research area will be beneficial in serving the community of researchers and practitioners by providing summarized information about Deep fake in all aspects, including available datasets, which are noticeably missing in previous surveys. Toward that end, we present a systematic literature review (SLR) on Deep fake detection in this paper. We aim to describe and analyze common grounds and the diversity of approaches in current practices on Deep fake

detection.

Our contributions are summarized as follows.

\_ We perform a comprehensive survey on existing literature in the Deep fake domain. We report current tools, techniques, and datasets for Deep fake detection-related research by posing some research questions.

\_ We introduce a taxonomy that classifies Deep fake detection techniques in four categories with an overview of different categories and related features, which is novel and the first of its kind.

\_ We conduct an in-depth analysis of the primary studies' experimental evidence. Also, we evaluate the performance of various Deep fake detection methods using

different measurement metrics.

\_ We highlight a few observations and deliver some guidelines on Deep fake detection that might help future research and practices in this spectrum.

The remainder of the paper is organized as follows: Section II presents the review procedure by defining interest research questions. In Section III, we thoroughly discuss the findings from different studies. Section IV summarizes

the overall observations of the study, and we present the challenges and limitations in Section V. Finally, Section VI concludes the paper.