DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

A Systematic Review of Deep Dream

Lafta R. Al-Khazraji[1], Ayad R. Abbas[2], Abeer S. Jamil[3]
[1,2] Department of Computer Science, University of Technology, Baghdad, Iraq
[1]General Directorate of Education of Salahuddin Governorate, Iraq
[3]Department of Computer Technology Engineering, Al-Mansour University College, Baghdad, Iraq
[1]l.alkhazraji@gmail.com, [2]ayad.r.abbas@uotechnology.edu.iq, [3]abeer.salim@muc.edu.iq

Abstract— *Deep Dream (DD) is a new technology that works as a creative image-editing approach by employing the representations of CNN to produce dreams-like images by taking the benefits of both Deep CNN and Inception to build the dream through layer-by-layer implementation. As the days go by, the DD becomes widely used in the artificial intelligence (AI) fields. This paper is the first systematic review of DD. We focused on the definition, importance, background, and applications of DD. Natural language processing (NLP), images, videos, and audio are the main fields in which DD is applied. We also discussed the main concepts of the DD, like transfer learning and Inception. We addressed the contributions, databases, and techniques that have been used to build the models, the limitations, and evaluation metrics for each one of the included research papers. Finally, some interesting recommendations have been listed to serve the researchers in the future.*

Index Terms— *Deep dream, deep CNN, gradient ascent, Inception, style transfer.*

I. INTRODUCTION

Deep Dream (DD) is a new technology presented in 2015 by Mordvintsev and his team at Google. Using CNN, DD aims to enhance image patterns with robust AI algorithms. Inception and Deep CNN represent the fundamental bases of a deep dream [1]. Google's Deep Dream is the typical application of this technique. DD works to improve the images by enhancing their visual attributes [2]. Visualization of DD has been used to determine whether the CNN correctly learned the right image features. So, the DD is created by increasingly feeding the image to the network where the first layers detect the first low-level features (i.e., edges). Then, the high-level features (i.e., faces and trees) appear, going deeper into the network. Finally, the rare final layers collect all those to configure combined effects (e.g., whole structures or trees) [3].

This study is the first systematic review of the Deep Dream. We collectedpublications that were relevant to DD by injecting a query containing the closest keywords related to DD in the databases of The .known publishing foundations like Springer, IEEE Xplore, MDPI, ScienceDirect, and others-well inclusion and exclusion criteria had been applied to preserve only the most relevant research and conference papers forsuch as transfer learning, ‹our study. The general concepts related to the DD .are addressed and explained clearly ‹and Inception ‹style transfer

The weakness in retaining human fluency, the sensitivity to the shape of an object, and the inability to generate objects with diverse topologies, in addition to the need for a huge dataset and the difficulty of modeling raw audio data, represent the major problems they faced in their studies.

This study's contributions start from being the DD's first review paper, addressing the reviewed papers' importance, challenges, limitations, and evaluation metrics used. Some recommendations are addressed to help researchers in the future, such as using a large dataset and exploiting the cloud applications to build a deep dream without the need for

Received 21/October/2022; Accepted 29/November/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

high-computing resources or wasting time. Also, combining DD with another technique, such as NST, can promote the resulting image and produce a stylized image.

II. BACKGROUND AND GENERAL CONCEPTS OF DD

This section shows the background of the Deep Dream. We start by introducing deep dreams and the situations that require applying them.

The deep dream is a computer vision program that uses the CNN algorithm [4] to transform images using motifs that the network has learned to recognize. Where it makes the mountaintop birds like beef birds and landscapes full of turtle-dogs and other chimeric creatures [5], applying DD in some situations is essential. It cannot be ignored in producing images [6], videos [7], [8], generating music [9], [10], NLP, and also for security [11].

Transfer learning is how a network's weights of the target model can be initialized by transferring it from a pre-trained model. Thus, avoiding the need for initialization and allowing the model to use prior learned weights [12]. So, transfer learning is a way of taking weights learned for a particular task and then reusing them for another similar one [3]. DD essentially depends on Deep CNN and Inception. Deep CNN is a distinct type of Neural Network that has excelled in several competitions involving Computer Vision and Image Processing. Deep CNN's excellent learning capabilities are attributed to utilizing the stages of feature extraction that can automatically learn the representations from data. The availability of vast data and advances in hardware technology have spurred CNN research, and fascinating deep CNN designs have lately been described [13]. While according to Arora et al. [14], Inception means building a layer-by-layer model in which the correlation statistics of the final layer must be analyzed and clustered into sets of units with high correlations. The units of the successor layers resulting from these clusters were connected to the units in the preceding layers. Christian Szegedy et al. [15] mentioned that every unit from the previous stage is supposed to correspond to a specific size of the input image. These units are organized into a set of bandpass filters. So, since the "Inception modules" are built on top of one another, the output correlation statistics would be changed. As higher layers collect elements of higher abstraction, the spatial density of these features is projected to diminish. The 33 to 55 convolutions ratio should grow [16].

Also, the significance of the structure of the inception layer came from the Hebbian precept of human learning, which states, "Neurons that fire together, wire together." So, it has been recommended that when someone builds the following layer in a DL model, he should consider the preceding layer's learnings. To explain this, it can be assumed that one of the layers in a DL model had been learned to target specific features of the face. The mesh's subsequent layer most likely concentrates on the image's public face to identify the various things there. So, to detect the various objects, the layer has to have suitable filter sizes. According to Yongcheng Jing et al. in their review, the Deep Dream was the first experiment to produce artistic images by reversing the representation of CNN with the techniques of IOB-IR. This was the base of neural style transfer [17].

Neural Style Transfer (NST) is a category of algorithms that permits us to use CNN to display the image contents in various styles [18]. Style Transfer can also be used in the text where the sentence style can be adjusted by rewriting the original style in a new style while keeping its semantic content [19].

In the content image, high-level features represent the objects and their arrangement, while the style image represents the image's texture, such as colors, sharpness, and styles. Thus, in a style image, the goal is to extract the style of the image, which is done by

Received 21/October/2022; Accepted 29/November/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

applying various filters to the style image and then taking the correlations between those filters. We are not concerned with the content image's information in this process. Only information about the style image is captured. So, the resulting image has been gathered from both the content and style image, where the contents of the former have been combined with the latter's style [20].

III. PROBLEM STATEMENTS AND DEEP DREAM IMPORTANCE

A. Problem statement

In bellow we summarize the problem of interest of this study:

Firstly, the increase in drug abuse cases and cases of schizophrenia and hallucinations for many people, especially young people, prompted researchers to develop a technique that helps specialists in psychiatric clinics and psychiatrists to see these hallucinations that their patients see to be able to improve the treatment methods used. Secondly, most of the images and art movements nowadays have become computer-dependent. Finally, researchers in artificial intelligence have always been looking for ways and techniques to show what is happening between the hidden layers.

B. Importance of the deep dream

DD has multiple usages:

- DD can be used as an art machine [21]
- DD modifies and decorates the images with motifs that the network has previously learned to distinguish [5].
- DD is used as a visualization tool to show what happens between the hidden layers of the CNN model [8].

IV. RESEARCH METHODOLOGY

This study introduced a systematic review of Deep Dream and its applications. This section presents the processes conducted to produce this systematic review. These steps are listed, starting with research questions and ending with eligibility criteria.

A. Research questions

The research questions and motivations helped us to get a better understanding of DD. All these are listed in Table I.

TABLE I. RESEARCH QUESTIONS AND MOTIVATIONS

Research questions	Motivations
RQ1) What is the benefit of the previous review of Deep Dream?	This is the first review of Deep Dream.
RQ2) What is the current status of the systematic literature review on the Deep Dream?	Deep Dream has been successfully implemented in various multimedia fields. The architecture of DD should be demonstrated.
RQ3) What is the distribution of research articles published on this topic based on the year of publication, authors, publishing house, and contributions?	Deep Dream has many challenges and difficulties. Also, it has a significant role in the image, video, audio, and even text processing.

Received 21/October/2022; Accepted 29/November/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

B. Research objectives

This research reviewed the most recent studies on the Deep Dream technique. The objectives of our review are:

- Identify the categories of the Deep Dream relevant studies.
- Identify the challenges of DD technology.

C. Data sources

This study was performed based on the research in the following electronic databases: Springer, MDPI, Elsevier, and IEEE Xplore. In addition, some other research papers were picked from ArXi, and Google Scholar, in addition to some cited references taken from the dependent websites. The research papers selected in this study were in various publication types, such as journal articles, conference papers, magazines, etc. *Fig. 1* shows the proportions of publications according to their source. *Fig. 2* shows the types of publications.

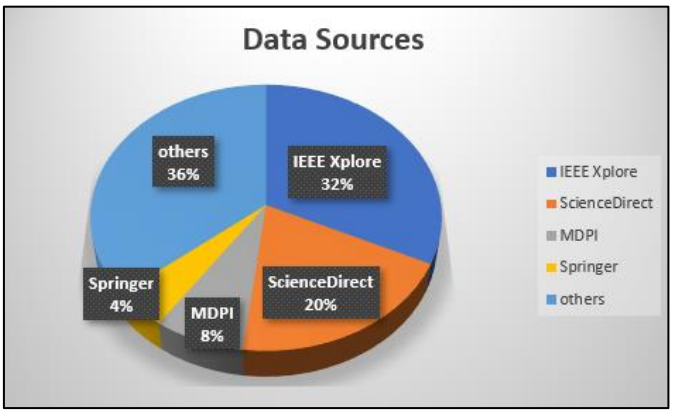

FIG. 1. THE PROPORTION OF PUBLICATIONS SOURCES.

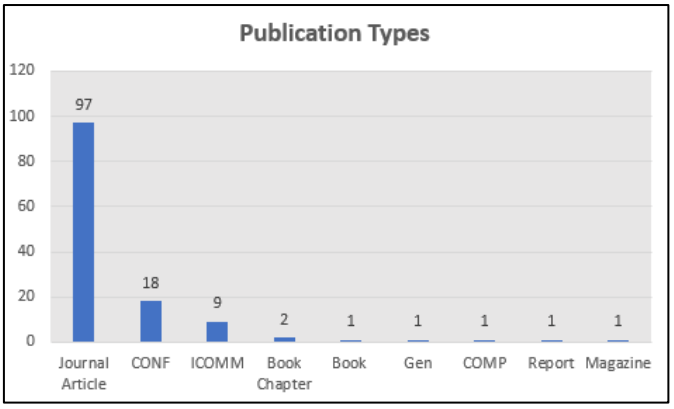

FIG. 2. THE PUBLICATION TYPES.

D. Keywords review search

This study was carried out by using specific keywords to accomplish the research questions and satisfy its objectives. This step was achieved on 1 June 2022 in Springer, IEEE, MDPI, Elsevier, and some books and websites. We used the query '("Deep Dream" OR "deep dream") AND ("hallucination" OR "hallucinated image" OR "text" OR "deep style" OR "image" OR "video" OR "audio")'. Using this query, we get the most recent search papers relevant to the topic of interest, which ranged from 2012 to 2022, as shown in *Fig. 3*.

Received 21/October/2022; Accepted 29/November/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

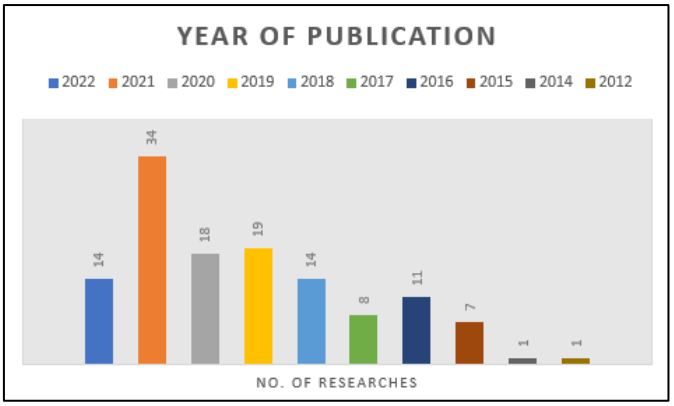

FIG. 3. THE COLLECTED PUBLICATIONS.

So, after we applied our research query, we did not find any review on the DD.

E. Study selection

Selecting relevant research is a complicated process, especially when various fields are taken into account. The first stage involved scanning titles and abstracts for duplicate and irrelevant research publications. This is the most critical stage, yet it's also the one that's often forgotten while researching a topic. The full-text reading of the chosen research papers was the second stage of this process.

F. Eligibility criteria

In this study, we gathered all articles and conference papers related to the subject of Deep Dream, and these papers were included in our study based on the criteria in Table II. After applying the keywords and bringing the results, the first stage is to remove the duplicated papers; then, we exclude the irrelevant papers. These processes are applied based on the title and abstract of those papers through the screening stage. The next stage is eligibility, where we assess the remaining research and works through taking the eligibility of the full text against the inclusion and exclusion criteria, which are:

- The works must be written in the English language.
- The works must be relevant and focus mainly on the Deep Dream technology.

Finally, only the remaining papers are included in this study which is only ten papers. Through searching the above keywords and applying the inclusion and exclusion criteria, we confirm that this study is the first review of the deep dream. *Fig. 4* shows the stages of select works of our study.

TABLE II. INCLUSION AND EXCLUSION ELIGIBILITY CRITERIA

Criteria	Specified Criteria
Inclusion	Research articles, conference papers, book sections, and some dependent websites.
Exclusion	• Non-English articles • Books, book sections and review deep learning papers. • Unrelated articles • Out-of-date publications

The remaining research papers are included, which have been explained in detail in the next section (Application of Deep Dream). We excluded the publications for several reasons, where the publications are either irrelevant (which are the papers that are unrelated to our subject) to our topic or out of date

Received 21/October/2022; Accepted 29/November/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

(articles that had been published before 2016), or excluded based on the background article (articles that are either a review deep learning papers or not related to computer science field). Finally, it may be the wrong publication type (report, website, program, etc.), as shown in *Fig. 5*.

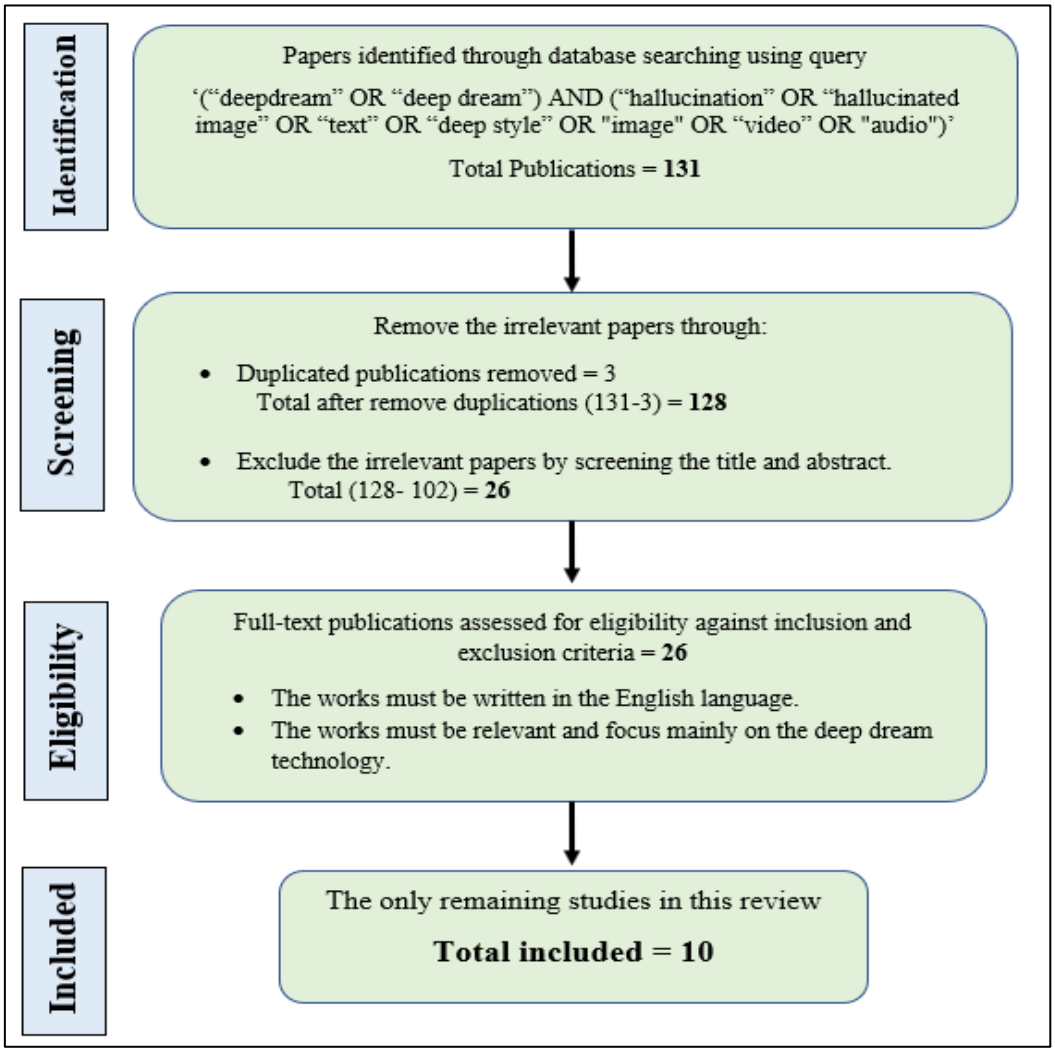

FIG. 4. FLOW DIAGRAM OF THE STUDY SELECTION, INCLUSION, AND EXCLUSION CRITERIA.

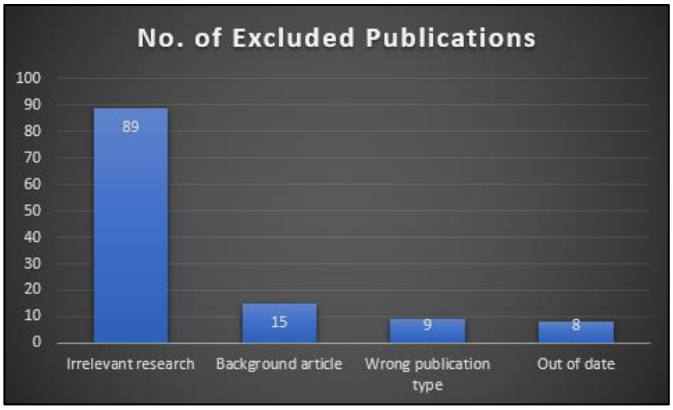

FIG. 5. THE EXCLUDED PUBLICATION PAPERS.

The investigation observed that DD had been applied to four multimedia elements: image, video, audio, and text. *Fig. 6* shows a diagram of the structure of our systematic review paper.

Received 21/October/2022; Accepted 29/November/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

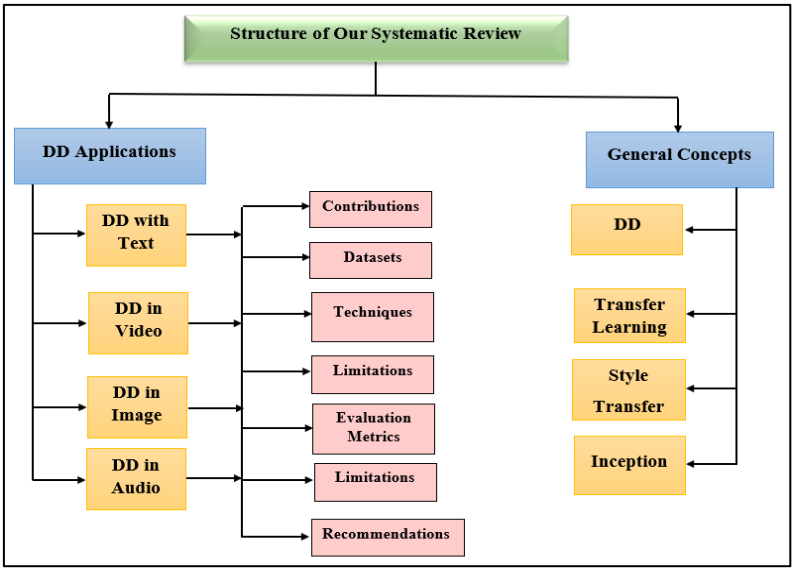

FIG. 6. DIAGRAM OF THE STRUCTURE OF THE SYSTEMATIC REVIEW PAPER.

V. APPLICATIONS OF THE DEEP DREAM

We start this section with Debora Pazetto Ferreira [23] approaching the intersecting issue of art and technology from the Flusserian's point of view applied to an intriguing example; images produced by a Google DD program. These surrealist-looking digital images result from a distortion in Google's artificial neural networks. He argued and claimed that these images pose a dilemma for philosophical dualisms such as those between AI and human intelligence, authorship versus anonymity, dominance, deviation, and art and technology. This section contains the included research papers that passed all the exclusion criteria, and it is divided into four parts according to the DD applications: NLP, image, video, and audio.

A. Deep Dream for NLP

Researchers started to use DD in NLP, especially with text. This process can be done by converting the words to embeddings, then achieving other text manipulation operations where the neural networks cannot process words directly. After that, DD can be used over text.

David Yue and Finsam Samson (2021) [22] presented a helpful technique called SleepTalk to understand better how the black box NLP algorithms make decisions. The SleepTalk technique enabled them to visualize learned representations of particular neurons in huge previously-trained networks.

Since DD cannot be used in NLP models immediately because of three critical differences between visual and textual data, these differences have been listed as follows: Firstly, visual images are represented in continuous data manifolds. On the other hand, words and sub-words are discretized tokens. Secondly, the interpretability of textual material depends on the complete preservation of semantic meaning and fluency, while visual representations are more tolerant of structural disturbances. Finally, NLP models often have sequential inductive prejudices, while computer vision models are inductively prejudiced toward locality; as a result, the auto-regressive character of NLP data necessitates semantic preservation at each time step.

So, to increase interpretability, they drew revelation from DD and activation maximization and applied it to NLP models. In their SleepTalk model, a neuron in the intermediate layer was chosen to be activated. The model's weights are fixed, and after that, the input embeddings are tuned using gradient ascent to improve the activity of that neuron.

Received 21/October/2022; Accepted 29/November/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

After finishing the optimization process, the input word embeddings are passed thru the network to the classifier, in which the words are transformed into an output text.

B. Deep dream for image

DD is most widely used for manipulating images, whether it is used for creating hallucination images [8], [23]–[25], or art images [26]. Many researchers employed the DD algorithm with deep learning techniques to produce images where the DD algorithm works to transform the strange landforms in one image (i.e., conical sandstone) into different shapes (i.e., faces, animals, and so on) [5].

Basma Abd El-Rahiem et al. (2022) [27] employed DD to present a multi-biometric cancellable scheme (MBCS) for creating cancellable fingerprint patterns that are both secure and effective, biometric modalities based on the veins of the finger and the iris of the eye. They harnessed the power of DL models to create a multi-exposure deep fusion module that generates a fused biometric template, which is then collected as the DD module's final cancellable template. They used Inception v3 as a pre-trained network. The loss value must be maximized during the gradient-ascent phase for DD to work. The goal of filter visualization is to maximize the value of a particular filter in a particular layer, which includes increasing the number of filter activations. Three key parameters control the gradient ascent process meters, which include: maximum loss $Lmax$, gradient step S, and several iterations I, which are applied to the loss gradients of the convent layers. Eventually, in every layer, a set of scales (i.e., octaves) produced by the dream rein junction identifies where the images would be processed. The rein junction can also upscale the input image at every layer, increasing its cancellability. As a result, each subsequent scaling is 1.4 times larger than the preceding one, resulting in a 40 percent increase in the image's starting dimensions. So, the re-injunction process begins with a small image gradually scaled up. Gradient ascent is used at each phase of the Deep Dream creation process, from the smallest image to the largest, to maximize the predetermined loss function. Furthermore, because the resulting image is upscaled by 40%, there is a requirement to reinject some of the lost info into the image after every consecutive scale-up (resulting in progressively fuzzy or pixelated images) to prevent losing much visual detail. The difference between the actual picture resized to size IL and the actual picture resized to a size that is quantified by the image details lost during transitioning from Is to IL once supplied a tiny image IS and a larger image IL.

Graeme McCaig et al. (2016) [24] presented two DD algorithms that achieved visual blending in CNNs; the first algorithm is Google Deep Dream [2], while the latter is the algorithm designed by (Gatys et al.), which took arbitrary images as input then they split and recombined its style and content through using neural networks to create artistic images [28], which are hence called deep style DS. The researchers applied GoogLeNet [16] for the DD and VGG [29] for the DS algorithm, the network trained on ImageNet and Cars datasets. Through their study of DD, they observed that two clichéd aspects deserve to stop and declared about them: firstly, the designed DD is a bottom-up discrimination network, and as a result, GoogLeNet ignores a significant amount of data about the tonic color of regions with retaining color contrast near edges. Secondly, the training data of ImageNet has a bias toward animals types as it represents a considerable part of the 1000 labeled classes, with a particular focus on the accurate distinction of dog breeds. So, tending to treat patterns as pertinent to dog features leads to the emergence of dog features. While car features emerge due to the network that has been trained on the car dataset.

They stated that DD and DS have difficulty replicating any training set images, and the output of DD/DS has visual similarity to the input images. Also, they find that DD tends to find similarity no matter whether it is more abstract or remote to the style of the guide image compared to the DS.

Hiroharu Kato et al. (2018) [30] proposed a model based on reconstructing a 3D mesh render the image. Besides that, they also fulfilled 2D-to-3D style transfer and 3D Deep Dream, where these

Received 21/October/2022; Accepted 29/November/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

applications illustrate the probability of integration of a mesh renderer into neural networks and the effectiveness of their proposed renderer. In DD, there are no lights in the rendered image. They used GoogleNet as a pre-trained model to train the DD model. They used Adam as a gradient descent.

Hongxu Yin et al. (2020) [31] proposed a new approach for synthesizing images from image distribution to train a deep neural network. Their approach is called DeepInversion. it consists of two parts; teacher and student logits. The teacher is the inverted trained network that starts from random noise without using any additional information on the training dataset. They used Deep Dream to build the DeepInversion by improving the image quality of DD through extending image regularization with a new feature distribution regularization term. Their model has been trained on CIFAR-10 and ImageNet datasets.

T. Tritva Jyothi Kiran (2021) [32] presented a computer vision algorithm called Deep Inceptionism learning or DD algorithm. In this algorithm, the training phase starts when an image enters the network, causing neurons to fire and generate activations. The principle of the DD algorithm is making some neurons fire more by modifying the input image (by boosting or activating neurons). DD algorithm provides the ability to choose particular neurons in distinct layers they are willing to, making them fire more conspicuously. This process would be repeated frequently until the input image contains all features needed by a particular layer. They continuously feed these images into the network, and the more they feed it into the network, the more they will be able to extract or see all of these strange elements in the actual image. So, their algorithm steps start by sending an image to a trained ANN, CNN, ResNet, etc. Then choose a layer (the top layer captures edges, while the deeper layers capture entire shapes like faces) and determine the activations (output) generated by the layer of interest and followed by calculating the activation gradient concerning the input image and altering the image to boost these activations, enhancing the patterns detected by the network, producing a trippy hallucinated image, and finally, repeating iteratively across multiple scales. They maximized the loss function by performing the gradient ascent at each layer. They added up all the losses from all the layers and passed them all, which was the same parameter plotting here or printing and returning it from the gradient ascent function. They could optimize their result by executing gradient ascent and running their dream algorithm.

C. Deep dream for video

Recently, DD has been widely spread in video activities, especially with virtual reality and simulation of visual hallucinations [7]. Therefore, we reviewed three publications that used DD with videos.

Antonino Greco et al. (2021) [33] presented the DD algorithm for creating visual stimuli that simulate hallucinatory states' perception to fix the difficulty of distinguishing the neurological consequences of psychedelic states from other physiological changes that have been brought on by drug consumption, where conducting tests on pharmacologically produced hallucinations is challenging, mainly due to ethical and legal concerns. They needed volunteers to achieve their studies. So, they brought participants. The volunteers were 20, 8 males and 12 females, and their average age was 26.4 in the 22−31.

Participants sat in a dimly lit booth, one meter from the CRT monitor. The original condition (OR) refers to a video clip retrieved from a movie. In contrast, the DD condition refers to a video clip modified from the original using DD. They chose to keep the requirements in the same order for two reasons. First, they learned through a small pilot that participants found exposure to the DD condition to be a "powerful" experience. They feared that there would be some tracing effects of DD over the OR condition by switching them. Second, many participants in both the pilot and the experiment did not

Received 21/October/2022; Accepted 29/November/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

recognize the DD movie as a modified version of the OR video. Therefore, they evaluated the likelihood of learning effects below.

The participants were shown a conventional movie first, then a modified one, while electroencephalography (EEG) was recorded. According to the findings, the frontal region had a higher entropy and lesser complexity during the edited movie than the regular one at different time scales. Furthermore, they discovered that the changed video triggered increased undirected connection and a higher level of entropy in functional connectivity networks. These data imply that DD and psychedelic substances cause similar altered brain patterns, demonstrating the utility of using this approach in neuroimaging studies to explore altered perceptual phenomenology.

They also expanded the stimulus set by exploring the DD parameter space by creating movies designed to match low-level to high-level activation layers consecutively. Because this study focused more on the semantic aspects of the stimulation, this would give information on how low-level features impacted brain dynamics. Furthermore, a thorough examination of how and to what extent DD can be compared to actual altered experience would significantly impact the area of psychedelic research, especially in light of recent results revealing that external stimulation plays a vital role in hallucinating state brain dynamics. Because DD, by definition, uses external stimulation to simulate hallucinogenic perception, it would be intriguing to employ it to investigate this element of psychedelic brain dynamics better. Finally, they showed how modern deep learning algorithms could provide psychedelic research with a new technique to explore altered perception using neuroimaging evidence.

D. Audio Deep Dream

In recent years, deep learning in audio signal processing has attracted much interest, and it is still a developing topic, so all that is reflexed in DD. This section presents the essential papers in which DD has been used with audio sound.

Diego Ardila et al. (2016) [34] took the first steps to apply the DD to audio by using raw audio to train a deep neural network (DNN) to implement a perceptual job. As a result, they trained a network to anticipate embeddings produced by a collaborative filtering approach. One significant difference is that they learn features directly from raw audio, resulting in a chain of distinguishable functions from raw audio to high-level features. The network is then used to extract "dreamed" audio instances using gradient descent. From 30-second snippets of music audio, they trained a network to predict 100-dimensional collaborative filtering (CF) track embeddings. The audio was sampled at $16kHz$, normalized by the maximum value of each mini-batch, and set to a mean of zero. The challenge here was that the scale of the embeddings differed significantly in this assignment, which can pose difficulties for the L2 loss function.

Furthermore, the popularity of any embedding is connected to its norm. As a result, they divided every embedding according to its L2 norm. They discovered that the learned first-layer filters still had much noise. Noise always have been presented in audio produced with this input and output. They started with noise or silence to produce sound, then optimized different targets inside the network. For example, maximizing the last layer's mean output could be a target. Then they employed gradient descent on the network's input. The result is not melodic, but it is also not pure noise.

Charis Cochran and Youngmoo Kim (2021) [35] explored feature visualizations that provide insight into learned models by applying DD to optimize inputs to activate specific nodes. They utilized a model that has been trained on the task of recognizing the most common instruments. They conducted two types of experiments: the first used a subset of the IRMAS data set that was evenly sampled from

Received 21/October/2022; Accepted 29/November/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

each of the 11 instrument classes. Then they utilized their DD algorithm implementation to construct two classes of modified test samples: one in which each modified spectrogram maximizes classification reliability for the proper instrument label and the other in which each spectrogram maximizes classification confidence for a randomly selected wrong instrument class.

After assessing network performance on these modified sets, they determined the Mean Squared Error (MSE) between the original and changed spectrograms and between properly and improperly optimized spectrograms. They also sonified various altered samples from both sets to compare to the original test samples. The second set of experiments included noise samples produced randomly and the same size as the input spectrograms. They treated these noise samples with the DD method to maximize the reliability of every instrument class. By using DD to modify the original test examples taken from the IRMAS data set, they could achieve extremes in network performance for every category, both positive and negative, because, in each "dreamed" example, the altered input was categorized with more than 90% confidence to the target category with 100% accuracy, regardless the altered input had been "dreamed" to the correct category or a randomly picked incorrect category. They then computed the MSE between the original and adjusted test sets and discovered that the altered samples varied from the original by only approximately 2 dB on average, resulting in nearly indistinguishable spectrograms. The adjustments performed by DD are still unnoticeable when listening to sonified samples and comparing them to the original test samples.

Additionally, they checked the MSE of correct and incorrect adjusted samples. They observed that for most instrument class pairs, the difference between these optimized samples was bigger than between the original and adjusted spectrogram. This demonstrates that the model is learning to distinguish between distinct instrument representations. The categories with lower MSEs may indicate that the instrument models are closer.

Halac F.N. and Delgadino M. (2021) [10] proposed a DreamSound, a creative sound-based adaption of DD taken from two methods; Sonification design and input manipulation. They introduced the original DD activation maximization function alongside three filter-based creative modifications. The three methods employed to design the model were the YAMNet model, Sonifying the gradients, and the DD function. The YAMNet model is a novel pre-trained deep network to classify sound [36]. In sonifying the gradients, they used a gradient ascent to maximize the loss function, where the gradient is known in this case as the 'gradient vector' between the loss and its corresponding input in a model. In this stage, they discovered similarities between their gradient sonifications and Herrmann's activation sonifications [37], which sounds like the noise filtered with dynamic spectral envelopes. Also, they discovered that the dynamic feature of these envelopes resembled the rhythmic parts of the input sound since they chose the last layer of YAMNet. The gradients' spectrogram appears to be a reversed version of the original sound's spectrogram. Finally, the DD function, where the gradients produced between the loss and the input are appended to the input and supplied back to the loop with appropriate attenuation. When a layer's activation is maximized, the outcome indicates a direction in the classifier's category space. In this scenario, the activation maximization is 'steered' in the direction of the original sound's class. A target is used to direct the model toward a different class. As a result, the sound would go to be more like itself or what the model deems to be its category. In conclusion, they found that listening to the audio files is a pretty idea.

Table III shows the contributions, the pre-trained models used, and the datasets we studied in our paper for each research paper.

Received 21/October/2022; Accepted 29/November/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

TABLE III. SUMMARIZES THE CONTRIBUTION, PRE-TRAINED MODEL, AND DATASETS USED FOR EACH DISCUSSED PAPER.

Study	Contribution	Pre-trained model	Dataset
[22]	Proposed a technique that helps for providing better interpretability of the NLP model.	Pre-trained BERT-Base-Uncased model	GYAFC dataset Sentences from the Wikipedia Corpus and the Book Corpus
[27]	Deploying a Deep Dream module to generate a cancellable template from the fused biometric image modalities.	Inception v3	Finger veins, where an infrared (IR) sensor had been used to collect it. Fingerprint image that is generated by using the LED of a particular device. CASIA Iris dataset.
[24]	Present a model that blends both DD and DS algorithms	• GoogLeNet for DD. • VGG for DS	CompCars database ImageNet
[32]	Using various pre-trained models (Inception, ResNet) to develop the DD model	• ResNet • Inception v3	ImageNet
[30]	They perform gradient-based 3D mesh editing operations, such as 2D-to-3D style transfer and 3D Deep Dream, with 2D supervision for the first time.	GoogleNet	ShapeNetCore dataset
[31]	Introduce the DeepInversion method for improving Deep Dream's image quality by extending image regularization with a new feature distribution regularization term.	• VGG • ResNet	• ImageNet • CIFAR10
[33]	used the Deep Dream algorithm to create visual stimuli that mimic the perception of hallucinatory states.	GoogleNet	• ImageNet • Places 365 dataset
[34]	Applying DD to audio	No pre-trained model. They built their CNN network.	They used 30-second clips of music audio as training data.
[35]	Apply the DD algorithm for exploring feature visualizations to get insight into the model.	No pre-trained model. They built their CNN network.	IR- MAS data set
[10]	Propose a DD called DreamSound, a creative adaptation of DD to sound addressed from input manipulation and sonification design.	YAMNet; is a pre-trained model for classifying sound.	Raw audio data was used.

As declared in Table III, ImageNet is the most repeatedly used dataset among all other datasets. *Fig. 7* shows the usage of all the datasets.

Received 21/October/2022; Accepted 29/November/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

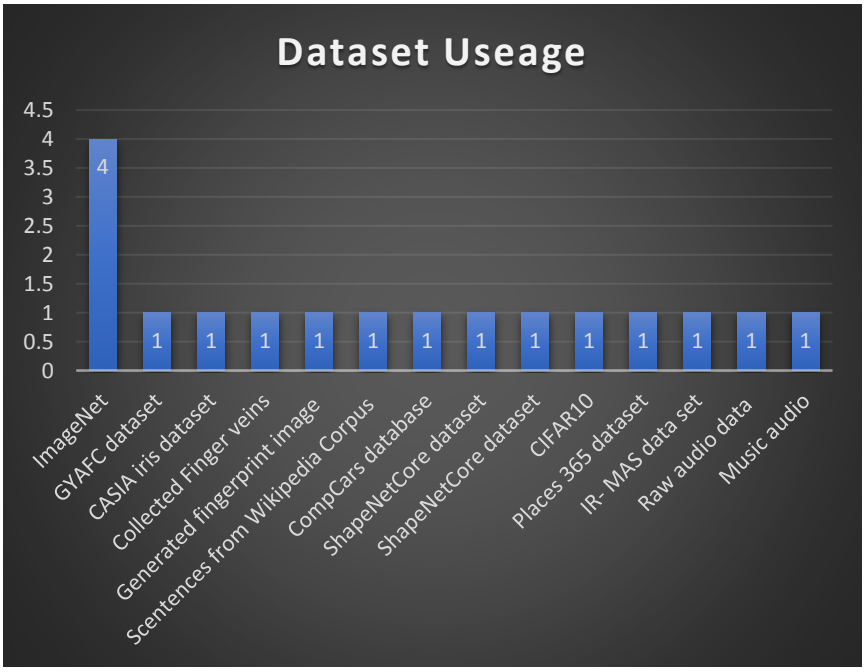

FIG. 7. THE USAGE OF EACH DATASET.

The pre-trained models that the research subject of the study used are shown in *Fig. 8*.

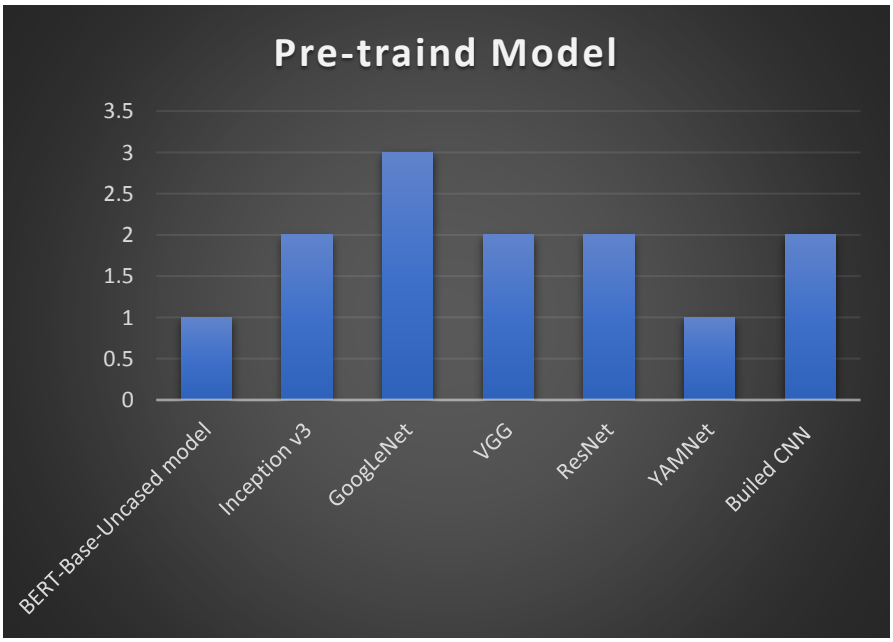

FIG. 8. THE NUMBER OF USAGES OF EACH PRE-TRAINED MODEL.

Fig. 9 shows the diagram of the DD applications and the techniques and datasets used for each application.

Received 21/October/2022; Accepted 29/November/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

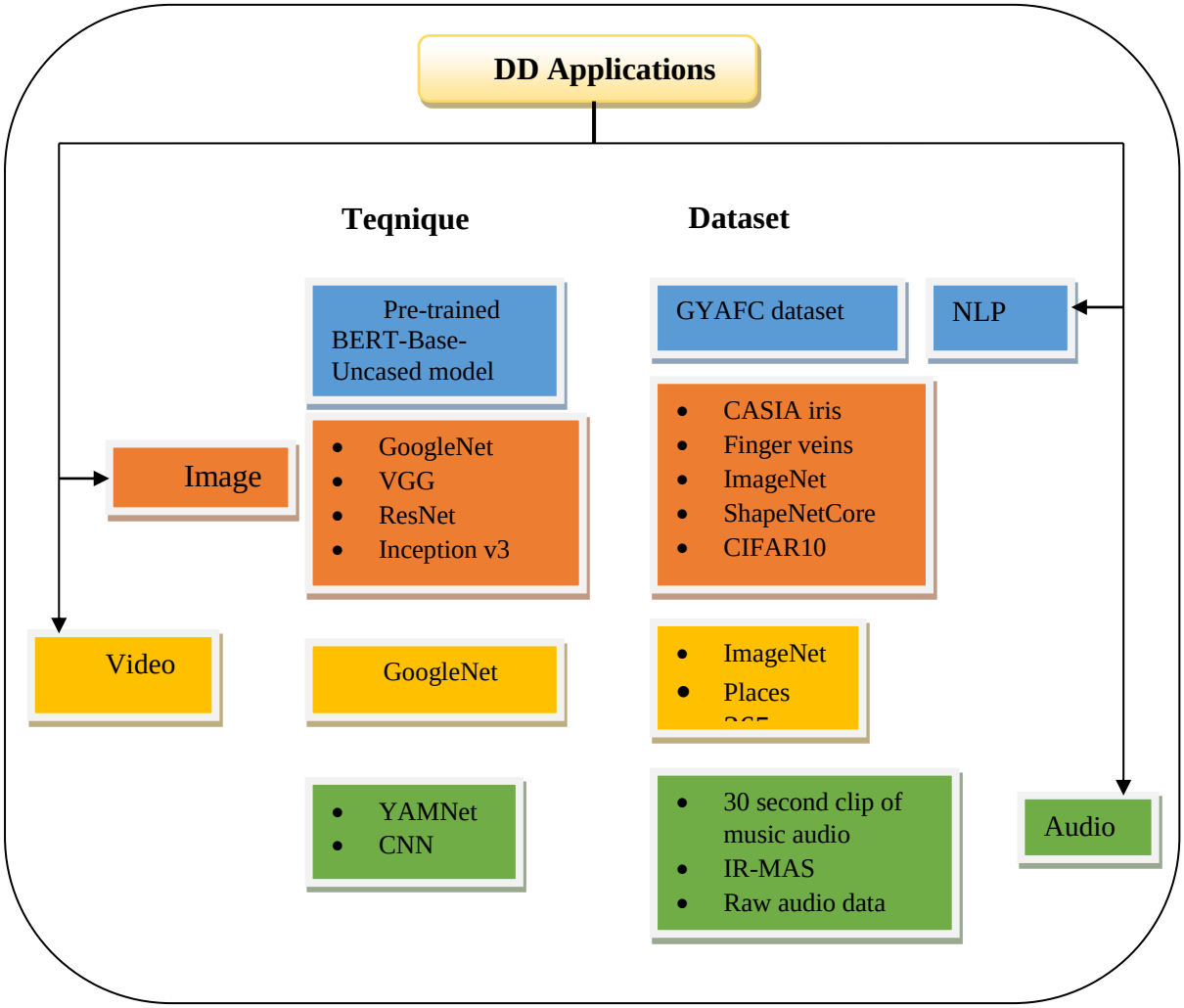

FIG. 9. DIAGRAM OF THE APPLICATIONS, TECHNIQUES, AND DATASETS OF EACH APPLICATION.

VI. LIMITATIONS AND PROBLEMS

The studies in our systematic literature review contributed to deep learning by employing DD to help solve various problems and provide excellent results. But, despite all the good results, there are still many limitations and problems in many of these studies; if we take [22], we find that although the authors performed excellent work, their work lacked retaining human fluency in the generated output. So, it requires the future to maintain fluency by perfectly interpreting the data. At the same time, the study of [30] has a clear disadvantage: its inability to generate objects with diverse topologies and very rough surfaces that could be generated sometimes in case of smoothness loss used. Also, their method is sensitive to an object's shape; it did not accomplish very well in the case of complicated shapes such as lamp, car, and table categories. In the [32], although the author implemented the DD algorithm and went inside it with a nice explanation of what exactly happened there, there was not any modification to the original algorithm and he just reimplemented the original DD algorithm. From the study of [24], combining hallucination with DD may result in unexpected visualization results, where the higher the typicality, the lower the novelty. Also, in [31], we observe that the authors themselves refer to limitations in their study, which include the fact that image synthesis requires an extended processing time even with high-performance hardware, and using default Gaussian distribution results in a high

Received 21/October/2022; Accepted 29/November/2022

ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

similarity of the color and background in the image. From [34], we conclude that according to Karan Goel et al. [38], it is very hard to model raw audio data because waveforms of audio have a high sample rate. The study of [27] used a minimal dataset in each modality (i.e., fingerprint, finger vein, and iris), each consisting of only nine images. Therefore, it was necessary to use a larger and more robust dataset.

VII. EVALUATION AND MEASUREMENTS

The studies that have been reviewed were evaluated and measured by using measurements that enabled the authors to evaluate the performance of their models. Below, we present some metrics and methods used to evaluate those studies.

We start with [33], where the authors used the entropy measure on EEG signal and, depending on a merging of several improvements of the traditional Shannon Entropy measure, which was entirely appropriate for EEG data. Here Permutation Entropy (PE) is used as a metric. For multi-time scales, multiscale weighted permutation entropy (MWPE) is used. And for the same purpose, complexity measure had been used depending on Jensen–Shannon complexity (JSC), which is a statistical measure that measures the signal's complexity relaying on the Jensen– Shannon divergence (JSD) produced, which measures the entropy and normalization constant of the signal.

The authors in [27] used two types of evaluations: visual and statistical. In the former, they used histogram and correlation analyses. Histograms provide insightful visualization for the distribution of pixel-wise intensity in the image. While in correlation analyses, a correlation coefficient metric is used to evaluate the relevance between a biometric image input and its produced cancellable template. At the same time, statistical evaluation has been divided into three metrics, which are quantitative, qualitative, and complexity analyses. Whereas in quantitative analysis, three metrics were used that named percentage pixel change rate (NPCR), unified average changing intensity (UACI), and peak signal-to-noise ratio (PSNR). in the qualitative analysis, two quality metrics were employed, which are spectral distribution (SD) and universal image quality index (UIQ). Finally, the complexity analysis was conducted by noting the required CPU operations.

In the study [22], the authors manually evaluated their study by assessing the fluency of the SleepTalk at various iterations to test the preservation of lucidity as their model (SleepTalk) progressed. Then, they took the dreamed words of the SleepTalk and studied its frequency to discover whether the model had the utmost affinities to particular types of tokens. Euclidian distance also had been used to measure contextual drift. The qualitative evaluation also had been used by visualizing the embedding at specific iterations using t-distributed stochastic neighbor embedding (t-sne). Additionally, the BLEU score was used as a measurement to understand the alteration in the semantic content.

In [31], the authors evaluated their work by using data-free pruning evaluation to get the best insight, and that was accomplished through studying: 1) a part of ImageNet (0.1 M), 2) 127k and 9.9 k unlabeled images from MS COCO and PASCAL VOC respectively. 3) they take 100k images generated from the BigGAN-deep model, and 4) the suggested method with the data-free setup. In addition to, various proposed new ways to evaluate their work.

In the study [24], the vector-similarity measure was used for deep style by combining it with the Gram matrix to maintain the source image's recognizable content. Regarding novelty, they suggested that distance measures generated from vector spaces are nearer to the visual similarity humans perceive than raw-pixel-based measures.

The authors of [30] evaluated their model in three ways; where they rebuilt a single 3D image and compared their model with a previous good model, which is a voxel-based model, then compared the accuracy of reconstruction of voxel-based and retrieval-based methods. The comparison depended on measures of the reconstruction accuracy of their model against the previous voxel-based. Then the

Received 21/October/2022; Accepted 29/November/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

evaluation was made qualitatively and quantitatively, wherein the qualitative evaluation focused on the shape, whereas the quantitative focused on the object's shape and position.

From the [35] study, the calculation of the Mean Squared Error (MSE) between the spectrograms (original and modified samples) and also between the optimized spectrograms (correctly and incorrectly) proved that their model learned the representation of the samples. Where the lower MSE, the closer the instrument models.

VIII. RECOMMENDATIONS

Our systematic review presents some recommendations to mitigate researchers' problems and limitations while doing their job.

There are many advantages of using Deep Dream, which can be used in many fields such as medicine, arts, security, and others. In some cases, Deep Dream may not be an ideal solution. So, we need to integrate Deep Dream with another technique, like style transfer. So, researchers in the field of Deep Dream need some improvements to solve the widespread problems and the challenges in that field.

Firstly, we start with the improvement of the dataset. The optimal performance of Deep Dream requires training the model on a large dataset. In the problems and limitations section, we noticed some problems because of using a small dataset, which causes the model not to generalize well.

Secondly, since implementing Deep Dream requires hardware with GPU and high RAM, which may increase the price of the computer device, it is recommended to turn to cloud platforms to implement those types of techniques; Google Colab is an example of this platform that offers a great programming environment with high computational resources.

Thirdly, it recommended giving more attention to employing Deep Dream in security, where it can be used for hiding particular details and providing more confidentiality.

Finally, we recommend concentrating more on integrating the Deep Dream technique with other techniques and thus can benefit from those hybrid techniques.

IX. CONCLUSIONS

Our study is the first systematic review of the Deep Dream, a deep learning technique found for the first time by the Google developer team. We depended on this systematic review of the research papers from strong publications' databases. In this study, we included most research related to the subject of Deep Dream based on our query after excluding the publications that did not meet the exclusion criteria. We have addressed the applications of Deep Dream: NLP, images, videos, and audio. Also, we focused on some of the essential operations with Deep Dream, like Inception, transfer learning, style transfer, and others. The contribution of each included research paper has been addressed in this study, and different studies used different databases to train their models, like ImageNet, CIFAR 10, Places 365 databases, etc. Also, the researchers used various techniques to build their models, such as GoogleNet, Inception V3, ResNet, etc. The evaluation metrics such as MSE, PE, JSC, and so on were used to check the model's performance.

We also included the limitations and problems with many of those research or conference papers, such as the strong hardware required and using a very limited dataset to train the models, although some limitations may not be addressed. We recommended some points to assist future researchers in doing the optimal work, saving time, and minimizing the required.

Received 21/October/2022; Accepted 29/November/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

REFERENCES

[1] A. Krizhevsky and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *In Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[2] A. Mordvintsev, C. Olah, M. Tyka, and E. Al., "Inceptionism: Going deeper into neural networks," *Google Research Blog*, 2015. [Online]. Available: http://googleresearch.blogspot.co.uk/2015/06/inceptionism-going-deeper-into-neural.html.

[3] M. A. Wani, F. A. Bhat, Khan, S. Afzal, A. Iqbal, and E. Al., "Basics of supervised deep learning," in *Advances in Deep Learning*, vol. 57, Springer Nature., 2020, pp. 13–30.

[4] K. Q. Li, "The Interpretation of (Deep) Dreams," in *International Symposium on Machine Learning and Art 2021, School of Creative Media, City University of Hong Kong*, 2021, no. June, pp. 179–180.

[5] B. Hayes, "Computer Vision and Computer Hallucinations," *American Scientist*, vol. 103, pp. 380–383, 2015.

[6] L. Berov and K. U. Kühnberger, "Visual hallucination for computational creation," in *Proceedings of the Seventh International Conference on Computational Creativity*, 2016, no. June, pp. 107–114.

[7] K. Suzuki, W. Roseboom, D. J. Schwartzman, and A. K. Seth, "A Deep-Dream Virtual Reality Platform for Studying Altered Perceptual Phenomenology," Springer US, 2017.

[8] K. Suzuki, W. Roseboom, D. J. Schwartzman, and A. K. Seth, "Hallucination machine: Simulating altered perceptual phenomenology with a deep-dream virtual reality platform," in *ALIFE 2018: The 2018 Conference on Artificial Life. MIT Press*, 2018, no. January, pp. 111–112, doi: 10.1162/isal_a_00029.

[9] J.-P. Briot, G. Hadjeres, Pachet, François-David, and E. Al., "Deep Learning Techniques for Music Generation - A Survey." arXiv preprint arXiv:1709.01620, 2017.

[10] D. Halac, F. N. C., & Matiás, "DREAMSOUND : DEEP ACTIVATION LAYER SONIFICATION," in *The 26th International Conference on Auditory Display (ICAD 2021)*, 2021, pp. 158–163, doi: doi.org/10.21785/icad2021.032.

[11] R. Arthi, A. R. Kishan, A. Abraham, and A. Sattenapalli, "Centralized Intelligent Authentication System Using Deep Learning with Deep Dream Image Algorithm," in *International Conference on Emerging Trends and Advances in Electrical Engineering and Renewable Energy*, 2021, pp. 169–178, doi: 10.1007/978-981-15-7504-4_18.

[12] L. R. Ali, S. A. Jebur, M. M. Jahefer, B. N. Shaker, and I. Technology, "Employing Transfer Learning for Diagnosing COVID-19 Disease," *Int. J. online Biomed. Eng.*, vol. 18, no. 15, pp. 1–12, 2022.

[13] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, 2020, doi: 10.1007/s10462-020-09825-6.

[14] S. Arora, A. Bhaskara, R. Ge, and T. Ma, "Provable bounds for learning some deep representations," in *31st International Conference on Machine Learning, ICML 14*, 2014, vol. 32, pp. 584–592.

[15] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *31st AAAI Conference on Artificial Intelligence (AAAI'17). AAAI Press*, 2017, pp. 4278–4284.

[16] C. Szegedy *et al.*, "Going Deeper with Convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9, doi: 10.1002/jctb.4820.

[17] Y. Jing, Y. Yang, Z. Feng, and J. Ye, "Neural Style Transfer : A Review," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 11, pp. 3365–3385, 2019.

[18] A. Singh, V. Jaiswal, G. Joshi, A. Sanjeeve, S. Gite, and K. Kotecha, "Neural Style Transfer: A Critical Review," *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3112996.

[19] M. Toshevska and S. Gievska, "A Review of Text Style Transfer using Deep Learning," *IEEE Trans. Artif. Intell.*, 2021, doi: 10.1109/tai.2021.3115992.

[20] H. Li, "A Literature Review of Neural Style Transfer," in *Princeton University Technical report, Princeton NJ, 085442019*, 2018.

[21] J. Zylinska, *AI Art. Machine Visions and Warped Dreams*, First Edit. OPEN HUMANITIES PRESS, 2020.

[22] D. Yue and F. Samson, "SleepTalk : Textual DeepDream for NLP Model Interpretability," *Stanford CS224N Natural Language Processing with Deep Learning*, 2021. [Online]. Available: https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1214/reports/final_reports/report041.pdf.

[23] S. Banerjee, W. J. Scheirer, K. W. Bowyer, and P. J. Flynn, "On hallucinating context and background pixels from a face mask using multi-scale GANs," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 300–309, doi: 10.1109/WACV45572.2020.9093568.

[24] G. McCaig, S. DiPaola, L. Gabora, and E. Al., "Deep convolutional networks as models of generalization and blending within visual creativity," in *Proceedings of the 7th International Conference on Computational Creativity, ICCC 2016*, 2016, pp. 156–163.

[25] D. Windridge, H. Svensson, and S. Thill, "On the utility of dreaming: A general model for how learning in artificial agents can benefit from data hallucination," *Adapt. Behav.*, vol. 29, no. 3, pp. 267–280, 2021, doi: 10.1177/1059712319896489.

[26] D. P. Ferreira, "Artificial dreams: Contemporary intersections between art and technology," *Polish J. Aesthet.*, vol.

Received 21/October/2022; Accepted 29/November/2022
 ISSN (Print) 1811-9212 ISSN (Online) 2617-3352

DOI: https://doi.org/10.33103/uot.ijccce.23.2.15

52, no. 1, pp. 41–55, 2019, doi: 10.19205/52.19.2.

[27] B. A. El-Rahiem, M. Amin, A. Sedik, F. E. A. El Samie, and A. M. Iliyasu, "An efficient multi-biometric cancellable biometric scheme based on deep fusion and deep dream," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 4, pp. 2177–2189, 2022, doi: 10.1007/s12652-021-03513-1.

[28] L. Gatys, A. Ecker, M. Bethge, and E. Al., "A Neural Algorithm of Artistic Style," *arXiv Prepr. arXiv1508.06576*, 2015, doi: 10.1167/16.12.326.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations. San Diego, CA, USA.*, 2015.

[30] H. Kato, Y. Ushiku, T. Harada, and E. Al., "Neural 3D Mesh Renderer," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3907–3916, doi: 10.1109/CVPR.2018.00411.

[31] H. Yin *et al.*, "Dreaming to distill: Data-free knowledge transfer via deepinversion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8715–8724, doi: 10.1109/CVPR42600.2020.00874.

[32] T. T. J. Kiran, "Deep Inceptionism learning performance analysis using TensorFlow with GPU – Deep Dream Algorithm," *J. Emerg. Technol. Innov. Res.*, vol. 8, no. 5, pp. 322–328, 2021.

[33] A. Greco, G. Gallitto, M. D'alessandro, and C. Rastelli, "Increased entropic brain dynamics during deepdream-induced altered perceptual phenomenology," *Entropy*, vol. 23, no. 7, pp. 1–12, 2021, doi: 10.3390/e23070839.

[34] D. Ardila, C. Resnick, A. Roberts, and D. Eck, "Audio Deepdream: Optimizing raw audio with convolutional networks.," in *Proceedings of the International Society for Music Information Retrieval Conference, New York, NY, USA.*, 2016, pp. 7–11.

[35] C. Cochran and Y. Kim, "Deepdream Applied To an Instrument Recognition Cnn," in *22nd Int. Society for Music Information Retrieval Conf.*, 2021.

[36] K. Drossos, S. I. Mimilakis, S. Gharib, Y. Li, and T. Virtanen, "Sound Event Detection with Depthwise Separable and Dilated Convolutions," in *Proceedings of the International Joint Conference on Neural Networks. IEEE.*, 2020, pp. 1–7, doi: 10.1109/IJCNN48605.2020.9207532.

[37] V. Herrmann, "Visualizing and sonifying how an artificial ear hears music," in *Proceedings of the NeurIPS 2019 Competition and Demonstration Track, vol. 123. PMLR*, 2020, pp. 192–202.

[38] K. Goel, A. Gu, C. Donahue, and C. Ré, "It's Raw! Audio Generation with State-Space Models," *ArXiv ID 2202.09729*, pp. 1–23, 2022.

Received 21/October/2022; Accepted 29/November/2022
 ISSN (Print) 1811-9212 *ISSN (Online) 2617-3352*