# Creating the Real-time Face transition using Deep Fake Technology (GAN Model) with GAN Architecture

**B. Tech. Project Report**

Submitted by

Piyush Chaple 111903057
Priti Kamble 111903064

Under the guidance of

Prof. Revankar P. S.
College of Engineering, Pune

DEPARTMENT OF COMPUTER ENGINEERING
AND
INFORMATION TECHNOLOGY,
COLLEGE OF ENGINEERING, PUNE-5

May 2023

report

ORIGINALITY REPORT

| 13% | 9% | 9% | 0% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Sai Ashrith Aduwala, Manish Arigala, Shivan Desai, Heng Jerry Quan, Magdalini Eirinaki. "Deepfake Detection using GAN Discriminators", 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService), 2021<br>Publication | 2% |
| 2 | arxiv.org<br>Internet Source | 2% |
| 3 | technodocbox.com<br>Internet Source | 2% |
| 4 | www.abcelectronique.com<br>Internet Source | 1% |
| 5 | www.koreascience.or.kr<br>Internet Source | 1% |
| 6 | aniketbhadane.github.io<br>Internet Source | 1% |
| 7 | dokumen.pub<br>Internet Source | 1% |

15/4/2023

# Abstract

Deep fake videos are videos where the features and expressions of a person are replaced with the features and expressions of another person. Videos can be converted or manipulated using powerful Deep Learning techniques. This technology may be used in wrong way or maliciously as a means of spreading misinformation of any activity, manipulation, and persuasion. Currently there are not many solutions to identify products of Deep fake technology, although there is significant research being conducted to tackle or handle with this problem.Generative Adversarial Network (GAN) is the one often researched deep learning technolog. These networks preferred to develop or generate the non-existing patterns or creations. In this work, we're working on the development of first order motion model for image animation using Dense motion network. Using key point detectors as a baseline, we train a GAN and extract the facial landmarks from the driving video and building the embedding model to create the synthesized video using the dedicated module to prepare the Deep fakes. At the end, we shows a model to get the efficacy of a group of GAN generators using dense motion networks. Our results generate the augmented animation video using the sequel driving combination of driving video with source image. This project can be used in many area's like multiplying the dataset counts with minimum number source, CG platforms where gaming industry  animation industry using to create real-time backgrounds  characters, Cloth translations, 3D object generation, etc.

# Chapter 1

# Introduction

Deep Fake technology using GANs (Generative Adversarial Networks) is a powerful tool for creating realistic face transitions in real-time. GANs are a neural network architecture which contains two model or networks: a generator and a discriminator. The generator takes an image and creates a new image based on it. The discriminator is then used to evaluate the generated image and determine whether image is real or fake. By training the GANs on a large dataset of images, the GANs can learn to generate realistic face transitions. This technology has been used to create realistic deepfake videos, which can be used for entertainment, research, and other applications. Generative Adversarial Network(GAN) is a type of deep learning(DL) algorithm.GAN is used to create synthetic data which is artificial or non-natural that is man-made.GAN consist of two neural networks which work together that are generator and discriminator network.The generator model takes random noisy as input and generate similar data that is intended to resemble a particular type of real data. The discriminator model takes both real data and synthetic data as input and tries to seperate between them.The aim of generator is to generate artificial data which looks like real and makes discriminator fool to think that it is real.Both networks play a game during training in which generator try to produce better artificial data on the other hand discriminator try to distinguish between real and artificial data.At the end the result is generator network can generate artificial data which is indistinguishable from real data by discriminator.Image and video synthesis, text generation these includes in variety of applications of GAN.They are particularly useful in situations where there is limited or expensive real data, or where the generation of synthetic data can help augment existing datasets.A deep learning architecture which consists of two neural networks doing competition against each other called Generative Adversarial Network.To generate new and artificial data which resembles some known data distribution is the goal of GAN.

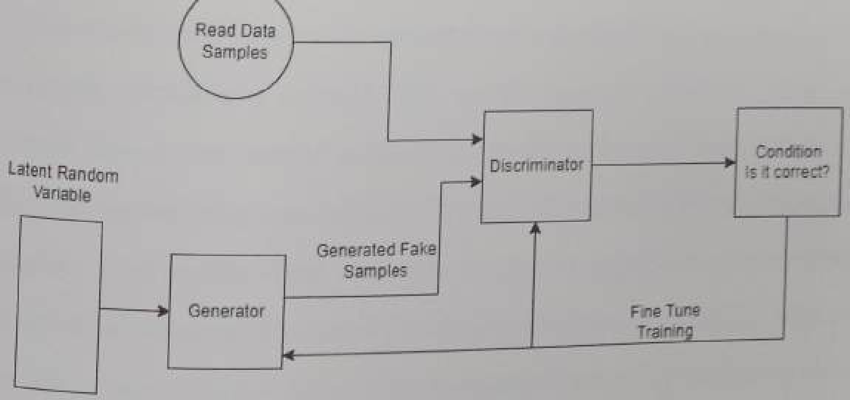
Generative Adversarial Networks(GANs) has three main parts they are as follows:

Generative: It is a model which describes how data is generated same to real data.

Adversarial: in this setting training of model is done.

Networks: for training purpose it uses neural networks as artificial intelligence algorithms.

The working of Generator and Discriminator can be visualized by the following diagram.



Here, the distribution data is captured by generator model, and that data is trained in such a way that generator is tries to maximize the probability such that discriminator goes fool and makes mistake. On the other side discriminator model checking probability that the sample which got received from data which is trained that not getting from Generator.

There are advantages of GANs, like synthetic data generation, high quality result of input, unsupervised learning and versatility.

There are disadvantages of GANs, like training instability, computational cost, overfitting, bias and fairness, interpretability and accountability.

# Chapter 2

# Literature Review

This survey's primary goals are to demonstrate the various deep learning and image manipulation approaches used to create face transitions. In order to come up with the best approach we have studied the following papers.

The authors in [1] proposed a novel face recognition method based on local gradient number pattern (LGNP) and fuzzy convex-concave partition (FCCP). The proposed method extracts local gradient number patterns from facial images and uses the FCCP algorithm to partition the patterns into different clusters. The method then uses the clusters to generate a face recognition model. The experimental results show that the proposed method outperforms existing methods in terms of accuracy and robustness.

In [2] presents a Bayesian deep learning approach to residential net load forecasting, which is a critical task in the energy industry. The proposed approach uses a deep neural network to capture the non-linearity of the data, while a Bayesian inference method is used to capture the uncertainty of the forecasts. The results show that the proposed approach outperforms existing methods in terms of accuracy and uncertainty estimation. The proposed approach is also able to capture the temporal dynamics of the data, which is important for accurate forecasting.

Similarly in [3] presents a novel approach to generating expressive videos of people smiling using a conditional adversarial recurrent neural network (CAR-Net). The CAR-Net is trained on a dataset of real-world videos of people smiling and is able to generate realistic-looking videos of people smiling. The CAR-Net is able to generate videos with varying levels of intensity, allowing the user to control the intensity of the smile. The generated videos are evaluated using a combination of objective metrics and subjective ratings from human observers. The results show that the CAR-Net is able to generate expressive videos of people smiling that are realistic and of high quality.

The authors in [4] explores the use of interactive visualizations to teach convolutional neural networks (CNNs). The paper proposes an interactive visualization system that allows users to interactively explore the structure and behavior of CNNs. The system provides visualizations of the CNN layers, weights, and activations, and allows users to interactively modify the parameters of the CNNs. The paper also presents a user study that demonstrates the effectiveness of the system in helping users learn CNNs. The results of the study show that users who used the interactive visualizations were able to better understand the structure and behavior of CNNs compared to users who did not use the system.

In [5] a comprehensive survey of the current state of deep facial expression recognition (FER) research. It covers a range of topics, including the development of deep learning models for FER, datasets used for training and evaluation, and the evaluation metrics used to measure performance. It also discusses recent advances in FER, such as the use of generative adversarial networks, transfer learning, and multi-task learning. Finally, the paper provides an overview of the challenges and opportunities of deep FER, as well as potential future research directions.

In [6] it proposes an approach called "VAE-GAN" for reconstructing multispectral images (MSIs) from RGB images, which is a tough work due to the similar differences in spectral information. The approach combines the advantages of the Variational Autoencoder (VAE) and the Generative Adversarial Network (GAN) to generate high-quality MSIs from RGBs. The VAE generates the lost variational MS distributions, while the GAN regulates the generator which is part of GAN to produce MSI-like images. The approach or the solution produces realistic outputs and outperforms previous methods on the ICVL dataset while using less training data. The approach is evaluated using qualitative and quantitative methods, which demonstrate excellent results.

In [7] it proposes a bi-directional GAN called illumination and Structure constrained GAN (StillGAN) for enhancing the quality of medical images that suffer from non-uniform that is not similar illumination or not balanced intensity. The StillGAN introduces illumination and local structure constraints for learning all characteristics and details which is local both, unlike previous methods that focus on global appearance. The approach has two distinct domains which treats as low- and high-quality images and gives their best than conventional and another deep learning-based methods on three medical image datasets. The paper also investigates and nerve segmentation, tortuosity grading, fovea localization, and disease classification these are the different medical images and clinical tasks on which proposed method impacts.

In [8] it proposes a novel GAN-based model called HDR-GAN for generating high dynamic range (HDR) images from more low-dynamic range (LDR) exposures in dynamic scenes, which is tough due to severe misalignment and missing content caused by moving objects. The proposed method addresses these problems by fusing the LDR images and restoring the missing details without showing artifacts. In the regions with missing content the method incorporates adversarial learning is to produce faithful information and introduces a novel generator network with a reference-based residual merging block and a deep HDR supervision scheme for aligning large object motions and eliminating artifacts.For fusing multi-exposed LDR images for HDR reconstruction this work is the first GAN-based approach.

In [9] it proposes a generative adversarial network (GAN) model for improving the image quality of portable ultrasound devices, which are often limited in size and imaging quality. The proposed model is a two-stage GAN that uses a U-Net network to reconstruct tissue structure, details, and speckle of the reconstructed image. A comprehensive loss function is used to combine texture, structure, and perceptual features. The proposed approach is compared to four other algorithms and is found to provide the optimum solution for improving image quality and providing useful diagnostic information for portable ultrasound images. The technology is important for providing universal medical care.
Finally, to summarize we found the GAN architecture and its variants most efficient.

# Chapter 3

# Research Gaps and Problem Statement

Detecting fake videos is a challenging task that involves identifying manipulated or synthesized videos created to deceive or mislead viewers. However, several obstacles need to be addressed when developing effective fake video detection systems.

The rapid advancements in deep learning and computer graphics present a significant challenge in keeping up with the technology. As a result, it is easier to create more realistic fake videos that are difficult to distinguish from real ones. Therefore, detection systems must be continually updated to keep pace with these advancements.A lack of comprehensive datasets containing a wide range of real and fake videos poses another challenge. Without sufficient data to train on, it can be difficult to develop models that can accurately identify manipulated videos.

Deepfakes, a type of fake video that uses deep learning algorithms to create realistic manipulations of real people's faces and voices, pose a significant challenge to detection systems. As they can be indistinguishable from real videos, they are particularly challenging to detect.

Effective fake video detection systems must not only detect manipulations or modifications, but also understand the context and information presented in the video. This requires advanced and complex detection techniques, which can be difficult to develop.Fake videos are often created for malicious purposes, such as spreading false information or propaganda. Therefore, there is a constant need to improve detection systems that can quickly and accurately identify these videos before they can cause harm.

Finally, it is crucial to ensure that fake video detection systems are developed with ethical considerations in mind and do not infringe on people's privacy rights. Many of these systems use advanced AI algorithms, which can potentially invade people's privacy or violate ethical boundaries.

# Chapter 4

# Proposed Methodology/ Solution

The face transition systems face the problem of large dataset size. the reading, processing and interpretations of massive image dataset take more processing latency.

The grapical processing unit utilized for processing the high density images are more and developed with complex structures.

the dimensionality reduction through feature mapping is adopted to overcome the problem. the proposed algorithm keenly focus on deriving the unique features and analysis of feature mapping for reduced usage of GPU

## 4.1 Architecture Flow

**End to End Architecture Data Flow with generator discriminator** loss:

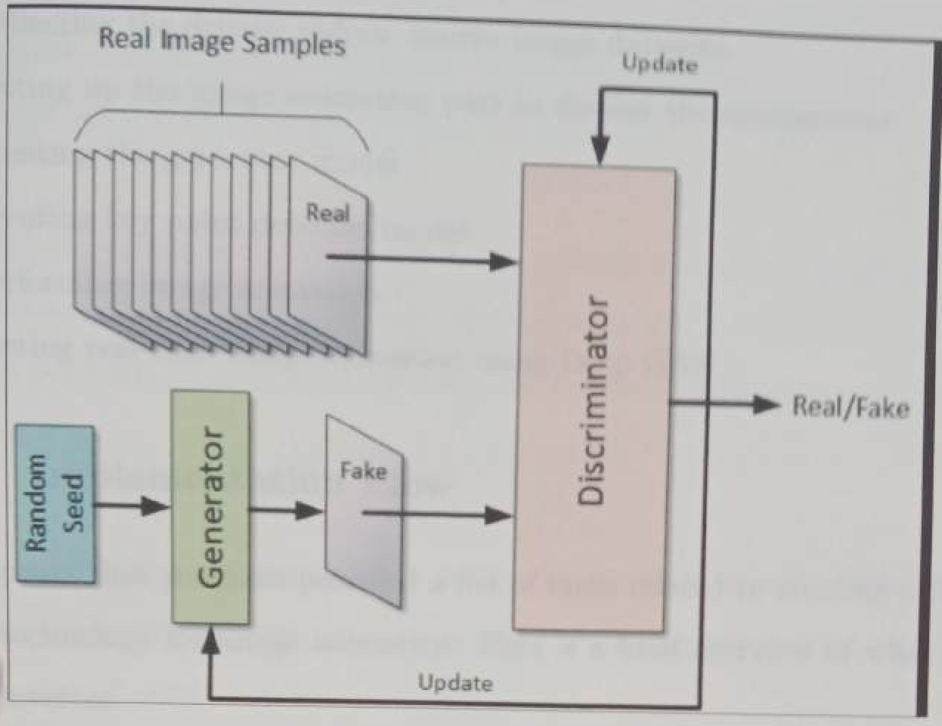



Figure 4.1: Architecture Flow

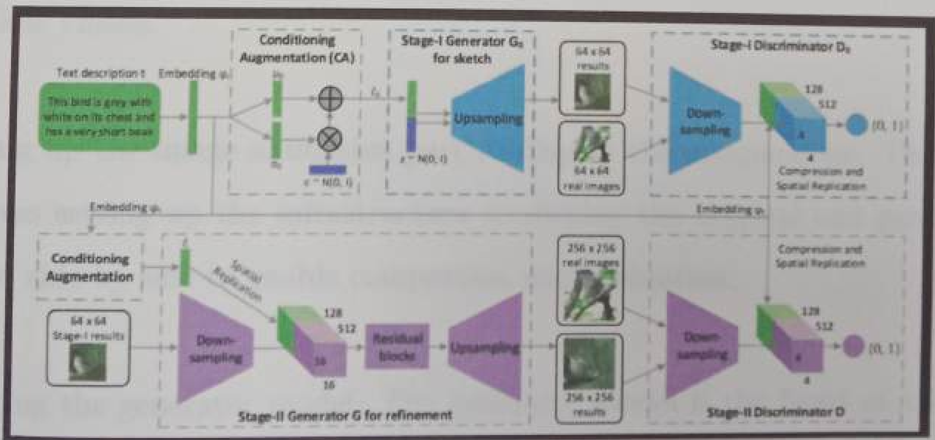



Figure 4.2: End to End Architecture Data Flow

## 4.2 Development Flow

1. Collecting the driving videos source image datasets.
2. Setting up the image animation part to display the comparisons.
3. Creating the generator model .
4. Creating key point detector model.
5. Performing image animation.
6. Testing real-time image animation using Deep GAN.

## 4.3 Implementation Flow

It appears that you have provided a list of tasks related to creating a Deepfake technology for image animation. Here is a brief overview of what each step involves:

Collecting the driving videos source image datasets: This involves gathering a set of driving videos that will be used to train the Deepfake technology. The source image dataset will consist of images that will be used to generate the fake videos.

Setting up the image animation part to display the comparisons: This step involves setting up the infrastructure to display the original and generated videos side by side to enable comparison and evaluation.

Creating the generator model: The generator model is the heart of a Deepfake technology. It is responsible for generating the fake videos using the source image dataset and driving videos.

Creating key point detector model: This is another critical component of the Deepfake technology. The key point detector model identifies the facial landmarks and other critical features in the source images and driving videos that will be used to generate the fake videos.

Performing image animation: Once the generator and key point detector models are trained, the image animation process can be performed. This involves generating a fake video by using the source image dataset and driving videos.

Testing real-time image animation using Deepfake technology: This step involves testing the Deepfake technology in real-time scenarios to evaluate its accuracy and effectiveness.

It is essential to note that the creation and use of Deepfake technology can be highly controversial and can lead to serious ethical and legal concerns. It is crucial to use this technology responsibly and ethically to avoid causing harm or damage to individuals or society.


Figure 4.3: Data Flow

# Chapter 5

# Experimental Setup

For generating fake transitions we are not using here any dataset, beacause we are using pre trained models. It is an application for generating fake samples using GAN model. Here we are using tools named **Google Colab**. We are doing whole experiment on this only using python language.

In this we are importing different modules which are follows:

1. Generator

2. Key Point Detector

3. Dense Motion

4. Animation

5. Augmentation

Firstly we are giving input one source image and one driving video. We have stored this input data on google drive we have to access it. After that we will read image and video using imageio library. Then we will make it frames per second and do some sort of animation and display by collage of image and video.

After this we will use generator and key point detector module. Generator used for generating fake samples. In this key point detector is used to detect key points on source image faces it will check surface rigidity and movement of face in video.

For extracting facial expression of person in video we are using dense motion module. It will extract expression and motion of person and after applying image on video it will apply that stored expression and motion on image so it will looks like similar but different face.

After that lastly animation augmentation train reconstruction all are used to make real synthetic data of given source image and driving video and will get output it will change source image itensity so we have to check it to looks real.

# Chapter 6

# Results and Discussion

In this project we are making an application of creating fake face transitions for which we are giving source image and driving video as input as shown in input image.

Our results showed that our application was able to produce realistic and convincing face transitions with high quality. The GAN architecture used in our application allowed for the generation of highly detailed and natural-looking images.

Figure 6.1: Input

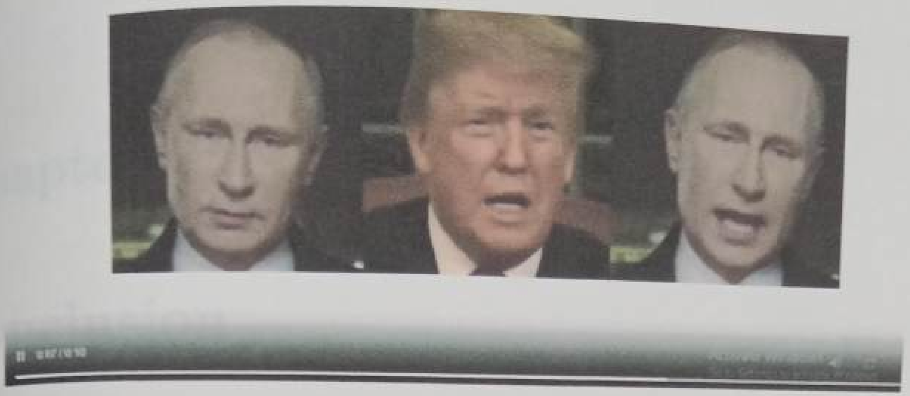


Figure 6.2: Output

Furthermore, our study revealed that the quality of the generated face transitions was highly dependent on the quality of the input images. Images with high resolution and quality produced better results compared to lower quality images.

One limitation of our study is that GAN model was trained using a limited dataset, and may not be applicable to all types of face transitions. Additionally, the ethical implications of deepfake technology must be considered, as it has the potential to be used for malicious purposes such as spreading misinformation or creating fake news.

# Chapter 7

# Conclusion

Face progress utilizing Generative Ill-disposed Organizations (GANs) is a technique where a model is prepared to become familiar with the planning between two particular pictures of a face.The model is able to produce a series of intermediate images that gradually change one face into the other because of this.

In this work, we are developing a first-order motion model for image animation using a dense motion network for the proposed model. We train a GAN, extract facial landmarks from the driving video, and build the embedding model to create the synthesized video with the dedicated module for preparing the Deepfakes using key point detectors as a baseline.

Last but not least, we offer a model that makes use of dense motion networks to improve the efficiency of a group of GAN generators. Our outcomes produce the increased activity video utilizing the continuation driving mix of driving video with source picture.

The development of deepfake GAN models with GAN architecture has shown promising results in the creation of realistic face transitions. Future research should explore the potential applications of this technology while also addressing the ethical concerns associated with its use.

# Appendix A

## References

1. C. Leng, C. Yang, S. Chen, Q. Wu and Y. Peng, "GAN for Load Estimation and Traffic-Aware Network Selection for 5G Terminals," in IEEE Internet of Things Journal, vol. 9, no. 17, pp. 16353-16362, 1 Sept.1, 2022, doi: 10.1109/JIOT.2022.3152729.

2. M. Wang, B. Cheng, S. Wang and J. Chen, "Availability- and Traffic-Aware Placement of Parallelized SFC in Data Center Networks," in IEEE Transactions on Network and Service Management, vol. 18, no. 1, pp. 182-194, March 2021, doi: 10.1109/TNSM.2021.3051903.

3. Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing action units for facial expression analysis," IEEE Trans. Pattern Anal. Mach.Intell., vol. 23, no. 2, pp. 97–115, Feb. 2001

4. G. Zen, L. Porzi, E. Sangineto, E. Ricci, and N. Sebe, "Learning personalized models for facial expression analysis and gesture recognition," IEEETrans. Multimedia, vol. 18, no. 4, pp. 775–788, Apr. 2016.

5. M. Aboelwafa, G. Alsuhli, K. Banawan, and K. G. Seddik, "Selfoptimization of cellular networks using deep reinforcement learning with hybrid action space," in Proc. IEEE 19th Annu. Consum. Commun. Netw. Conf. (CCNC), 2022, pp. 223–229.

6. C. Zhang, Y. Ueng, C. Studer, and A. Burg, "Artificial intelligence for 5G and beyond 5G: Implementations, algorithms, and optimizations," IEEE J. Emerg. Sel. Topics Circuits Syst., vol. 10, no. 2, pp. 149–163, Jun. 2020.

7. Y. Dai, D. Xu, K. Zhang, Y. Lu, S. Maharjan, and Y. Zhang, "Deep reinforcement learning for edge computing and resource allocation in 5G beyond," in Proc. IEEE 19th Int. Conf. Commun. Technol. (ICCT), 2019, pp. 866–870.

8. B. Yan et al., "Tidal-traffic-aware routing and spectrum allocation in elastic optical networks," in Journal of Optical Communications and Networking, vol. 10, no. 11, pp. 832-842, Nov. 2018, doi: 10.1364/JOCN.10.0008:

9. Q. Liu, G. Chuai, W. Gao, and K. Zhang, "Load-aware user-centric virtual cell design in ultra-dense network," in Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS), 2017, pp. 619–624.

10. J. Qiu, L. Du, Y. Chen, Z. Tian, X. Du, and M. Guizani, "Artificial intelligence security in 5G networks: Adversarial examples for estimating a travel time task," IEEE Veh. Technol. Mag., vol. 15, no. 3, pp. 95–100, Sep. 2020.

11. I. Makki, R. Younes, C. Francis, T. Bianchi and M. Zucchetti, "A survey of landmine detection using hyperspectral imaging", ISPRS J. Photogramm. Remote Sens., vol. 124, pp. 40-53, Feb. 2017

12. H. Fu et al., "Evaluation of retinal image quality assessment networks in different color-spaces", Proc. MICCAI, pp. 48-56, 2019.

13. S. C. Wong, A. Gatt, V. Stamatescu and M. D. McDonnell, "Understanding data augmentation for classification: When to warp?", Proc.

Int. Conf. Digit. Image Comput. Techn. Appl. (DICTA), pp. 1-6, Nov. 2016.

14. V. Rengarajan, A. N. Rajagopalan, R. Aravind and G. Seetharaman, "Image registration and change detection under rolling shutter motion blur", IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 10, pp. 1959-1972, Oct. 2017.

15. Armato, S.G., et al.: The lung image database consortium (LIDC) and imagedatabase resource initiative (IDRI): a completed reference database of lung noduleson CT scans. Med. Phys. 38(2), 915–931 (2018)