Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 18 No. 1 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijim.v18i01.42673

PAPER

A Visual Computing Unified Application Using Deep Learning and Computer Vision Techniques

Sowmya B.J.[1], Meeradevi[2], S. Seema[3], Dayananda P.[4](✉), Supreeth S.[5], Shruthi G.[5], S. Rohith[6]

[1]Department of Artificial Intelligence and Data Science, Ramaiah Institute of Technology, Bengaluru, Karnataka, India

[2]Department of Artificial Intelligence and Machine Learning, Ramaiah Institute of Technology, Bengaluru, Karnataka, India

[3]Department of Computer Science and Engineering, M S Ramaiah Institute of Technology, Bengaluru, Karnataka, India

[4]Department of Information Technology, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, Karnataka, India

[5]School of Computer Science and Engineering, REVA University, Bengaluru, Karnataka, India

[6]Department of Electronics & Communication Engineering, Nagarjuna College of Engineering & Technology, Bengaluru, Karnataka, India

dayananda.p@manipal.edu

ABSTRACT

Vision Studio aims to utilize a diverse range of modern deep learning and computer vision principles and techniques to provide a broad array of functionalities in image and video processing. Deep learning is a distinct class of machine learning algorithms that utilize multiple layers to gradually extract more advanced features from raw input. This is beneficial when using a matrix as input for pixels in a photo or frames in a video. Computer vision is a field of artificial intelligence that teaches computers to interpret and comprehend the visual domain. The main functions implemented include deepfake creation, digital ageing (de-ageing), image animation, and deepfake detection. Deepfake creation allows users to utilize deep learning methods, particularly autoencoders, to overlay source images onto a target video. This creates a video of the source person imitating or saying things that the target person does. Digital aging utilizes generative adversarial networks (GANs) to digitally simulate the aging process of an individual. Image animation utilizes first-order motion models to create highly realistic animations from a source image and driving video. Deepfake detection is achieved by using advanced and highly efficient convolutional neural networks (CNNs), primarily employing the EfficientNet family of models.

KEYWORDS

deep learning, convolution neural networks (CNNs), computer vision techniques, visual computing

1 INTRODUCTION

Vision Studio is a unified application that applies deep learning and computer vision techniques to visual computing. The application aims to utilize a diverse range of modern deep learning and computer vision principles and techniques to provide a broad array of functions in image and video processing.

The computer science disciplines that deal with various visualization techniques, such as image processing, computer vision, virtual and augmented reality, and

Sowmya, B.J., Meeradevi, Seema, S., Dayananda, P., Supreeth, S., Shruthi, G., Rohith, S. (2024). A Visual Computing Unified Application Using Deep Learning and Computer Vision Techniques. *International Journal of Interactive Mobile Technologies (iJIM)*, 18(1), pp. 59–74. https://doi.org/10.3991/ijim.v18i01.42673

Article submitted 2023-06-29. Revision uploaded 2023-10-19. Final acceptance 2023-10-24.

video processing, are commonly grouped together. Deep learning techniques are used to construct a more abstract and comprehensive representation of images and videos. Deep learning is a distinct class of machine learning techniques that utilize multiple layers to progressively extract more complex features from raw input. This is beneficial when using a matrix as input for pixels in a photo or frames in a video. By applying the previously described techniques, an individual can replace a person's presence in a video or image with the likeness of someone else. Deepfakes are a type of manipulation or generation of visual and audio content that utilizes powerful deep-learning strategies. They have a high potential to deceive. Therefore, methods must be developed through which individuals can detect deepfakes and protect themselves from misinformation. Computer vision is a field of artificial intelligence that teaches computers to interpret and understand the visual world. Utilizing advanced algorithms from cameras and videos, as well as deep learning models, machines can accurately recognize and classify objects, enabling them to respond to visual stimuli. The application aims to utilize a diverse range of modern deep learning and computer vision principles and techniques to provide a wide array of functions in image and video processing [1].

The main motivation of Vision Studio is not only to enable users to utilize powerful techniques for safely manipulating media and understanding the underlying technology, but also to raise awareness about these techniques and the potential outcomes that may arise. For example, misinformation or fake news can give rise to various problems in society. Giving the general public an idea of the far-reaching power of such techniques is the first step towards combating misinformation and fake news and ultimately creating a better society [2].

The main contributions that have been implemented are as follows:

- Deepfake creation: Users input source images and target videos, and algorithms superimpose the source face images onto the target videos, creating fake videos. This allows for editing without having to reshoot dialogues. Digital watermarks aid in identifying these counterfeits, reducing malicious intent.
- Digital aging/De-aging: The computer alters an individual's appearance across different ages while preserving their identity by modifying their head shape and texture while also maintaining key facial attributes.
- Bringing images to life: Digitally, first-order motion models can be used to transform an individual in a photograph into a realistic animation, resembling the images inspired by Harry Potter.
- Deepfake detection: The process utilizes multiple identification methods to detect deepfakes, incorporating countermeasures such as confirmation and provenance innovation, in addition to media proficiency measures.

All of the aforementioned functions will be provided in a single, user-friendly application that is easy to use. This will abstract the underlying technology and allow any person, regardless of their familiarity with it, to fully utilize the application. An application that can perform the aforementioned four objectives, as described above, i.e.,

- Deepfake creation
- Digital aging/De-aging
- Bringing images to life
- Deepfake detection

The user can utilize these features according to their preferences and delve into the realm of deep learning and computer vision. This application can deliver the four objectives mentioned above, helping individuals gain a better understanding of the current application of deep learning and computer vision methods that impact their daily lives. Individuals will be able to distinguish between authentic and manipulated media. Adding more applications of deep learning methods will benefit both the general public and companies.

2 LITERATURE SURVEY

Visual computing is an interdisciplinary field that encompasses computer vision, computer graphics, and image processing. It focuses on the acquisition, analysis, manipulation, and synthesis of visual data, making it a fundamental component of modern technology and entertainment [3]. The AI generation, driven by advancements in deep learning and neural networks, has revolutionized the production of visual content. Machine learning models can generate images, videos, and other visual media, making it possible to automate content generation and manipulation [4]. Deepfakes are a prominent application of AI generation within the field of visual computing. They refer to the synthesis of highly convincing, yet entirely fabricated, visual content, with a specific focus on face swapping and face re-enactment [5]. Deepfakes have attracted significant attention due to their potential for both creative applications and misuse, raising concerns about their ethical and legal implications. In the early stages of deepfake development, traditional methods were commonly used. These approaches involved face detection in the original image, selecting a suitable face image from a candidate set, replacing the original image's facial features, and adjusting for lighting and color to match the original scene [6]. Recent advancements have introduced more sophisticated deepfake techniques. Modern approaches employ two encoder-decoder pairs. The encoder extracts latent features from the original image's face, while the decoder reconstructs the face using these latent features. By exchanging decoders, a unique encoder from the source picture and a decoder from the target picture are utilized to recreate the target image, preserving the features of the source image and producing more convincing and realistic results [7].

2.1 Deepfake creation

Deepfake creation is an application that primarily focuses on visual deepfake techniques, specifically face swapping and face re-enactment. In the past, traditional methods [8–9] were employed to achieve deepfake creation [26–32]. The process includes detecting the face in the original image, selecting a suitable face image from a candidate set, replacing the facial features of the original image with the selected face, and adjusting for lighting and color to match the original scene.

However, recent advancements in deepfake technology have led to the utilization of more sophisticated methods. Modern approaches utilize a different technique: two encoder-decoder pairs are employed. The encoder extracts latent features from the original image's face, while the decoder reconstructs the face using these latent features. After preparation, the decoders are swapped. A unique encoder is used to

encode the source picture, and a decoder is used to recreate the target picture, incorporating the features of the source picture [10–11].

2.2 Digital Ageing/De-Ageing

Facial age transformation plays a crucial role in cross-age recognition and various entertainment-related applications. Several strategies have been developed to address age recognition and age modification in facial images. This literature review explores key techniques in this field. One of the approaches discussed is face-ageing with identity-preserved conditional generative adversarial networks (IPCGANs) [12]. This method utilizes IPCGANs to perform age transformations on facial images. Another method mentioned is lifespan-age transformation synthesis [13]. This technique utilizes an identity encoder to extract attributes that are pertinent to a person's identification from the input image. This allows for age transformation while maintaining the integrity of the person's identity information. High-resolution face-age editing [14] is also explored in this review. This technique utilizes an encoder-decoder architecture to generate a latent space that represents facial characteristics, along with a component regulation layer that enables accurate age modification of individuals in the image. Lastly, the review discusses the method titled, "only, a matter of style: age transformation using a style-based regression model" [15]. In this approach, a pre-trained unconditional GAN is used to encode real facial images directly into a latent space while incorporating an aging shift to achieve age transformation.

2.3 Bringing images to life

Image animation is a dynamic process that involves using computer software to create videos. It combines features extracted from a source image with motion derived from a driving video. Over time, various methods and technologies have been utilized to accomplish image animation. Historically, image animation primarily relied on traditional, object-specific methods. However, with the advancement of technology, more modern techniques have emerged, including the use of generative adversarial networks (GANs) [16] and variational auto-encoders (VAEs) [17–18]. Monkey-Net [19] is a novel and innovative model for picture animation. It encodes movement data using central points that are learned in a self-directed manner, allowing for more dynamic and creative animations. First-order motion models [20–21] have proven to be effective, even when dealing with significant changes in object pose. It achieves this through self-learned focal points, consideration of local contextual changes, a generator that is aware of potential limitations, and the use of equivariance loss to improve the evaluation of local contextual changes.

2.4 Deepfake detection

This text highlights the significance of various multimedia forensic techniques designed to detect deep-fake content. These techniques encompass a range of

approaches, including data-driven classification, GAN fingerprints, video frame synchronization, anomaly detection, and more [22–23]. For image forgery detection, a feature extractor was trained using support vector machines (SVM) for classification. This technique is particularly effective in identifying manipulated images [24]. Similar techniques used for image forgery were employed for video deepfake detection. Frame-by-frame analysis is used to identify changes made in deep-fake videos. To enhance the ability to detect various details in the video format, improvements were introduced. This includes distinguishing facial features such as lip, head, and hand movements [25]. Audio detection is an essential component of deepfake detection. Techniques such as high-frequency cepstral coefficients (HFCC) and constant-Q cepstral coefficients (CQCC) were used to analyze audio content. SVMs are used to recognize various sound patterns. To combat replay attacks, the transmission line cochlea-sufficiency tweak (TLC-AM) and TLC-recurrence balance were employed to train a Gaussian mixture model (GMM). This helps in identifying and countering attempts to imitate authentic audio or video content.

Deep learning and computer vision have experienced exponential growth in the last couple of years, and their applications are leading to daily transformations. Some cutting-edge developments have been discussed in the survey. However, it lacks specific examples of modern deep-fake techniques, making it challenging to assess their effectiveness and limitations. It does not delve into potential ethical concerns or issues surrounding deepfake technology. This does not provide details on their accuracy, limitations, or any comparative analysis. To gain a deeper understanding of the current state of age transformation techniques, it is necessary to conduct a more comprehensive analysis of their strengths and weaknesses. This text does not discuss the ethical concerns related to deepfake technology or its potential real-world applications and implications.

3 DESIGN AND IMPLEMENTATION

The work consists of achieving four objectives: deepfake creation, digital ageing and de-ageing, bringing images to life, and deepfake detection. This section presents the design and architecture used to implement this work. Figure 1 illustrates the comprehensive flow of the backend operations. The end users are exposed to a frontend GUI, as shown in Figure 2, which connects to an internal gateway in the application. All content is served using a Flask backend.

There are three separate modules: the Deepfake module, the Digital Aging module, and the Image Animation module. The Deepfake module consists of two sub-modules: the Creation module and the Detection module. The Creation module primarily handles the task of creating deepfakes, while the Detection module is responsible for detecting deepfakes. The Digital Aging module and Image Animation module handle digital aging and bringing images to life, respectively. Each module is connected to the backend server, which houses the trained model and also stores the input for future training, which takes place in the training module.

Each module is connected to an endpoint. When the user input is received, the module processes it and returns the processed output. The training module aims to facilitate continuous improvement in the model iteratively by utilizing newer data stored on the server to create more robust models.

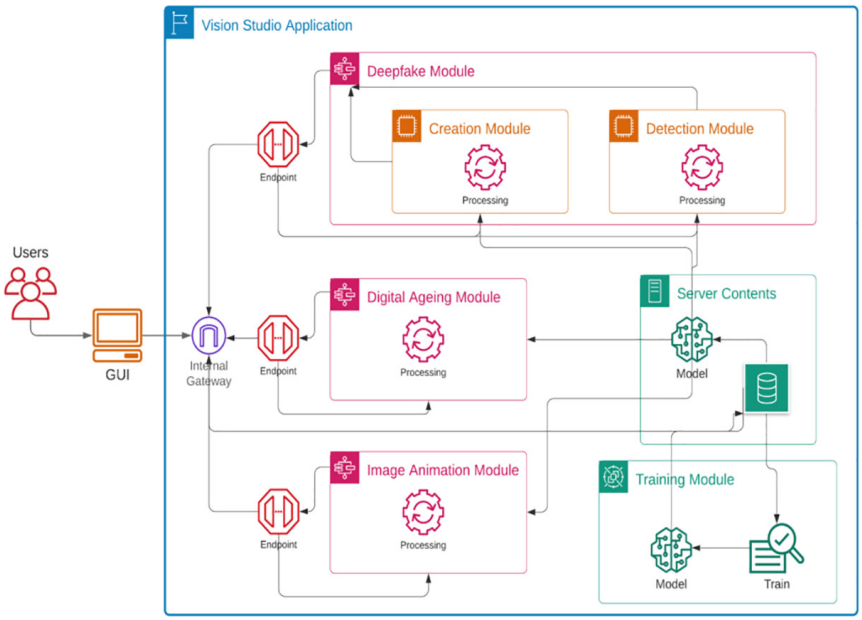

Fig. 1. Proposed system architecture

3.1 Deepfake module

The Deepfake module consists of two sub-modules: Deepfake Creation and Deepfake Detection, each with its own models. The Deepfake Creation module allows users to create deepfakes by providing image and video inputs and receiving a video output. The Deepfake Detection module enables users to detect deepfakes by inputting a video and receiving an output probability score. This score indicates the likelihood of the input being a deepfake.

Figure 2 represents a deepfake creation model that utilizes two encoder-decoder pairs. Two organizations use the same encoder and different decoders for the training process (top). An image of face A is encoded using the shared encoder and decoded with decoder B to create a deepfake (bottom). The application implements and aims to enhance the most commonly used pipeline for creating deepfakes as of now.

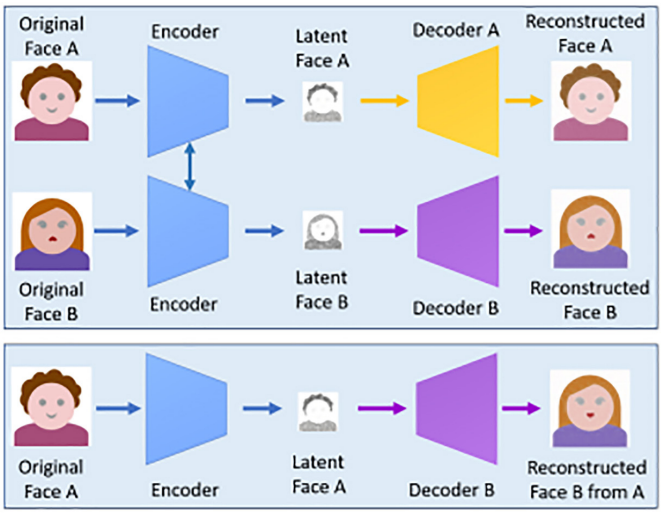

Fig. 2. Explanation of deepfake creation

3.2 Digital ageing module

The Digital Aging module consists of two separate models: one for male faces and one for female faces. Each model takes an image as input and provides a video output, along with image outputs for a specific age range that is being considered.

An individual's appearance changes as they age, while their personality remains intact. To perform these tasks, the computer needs to modify the shape and texture of the head while maintaining the essential facial features of the input face.

With the development of GANs, it is now possible to synthesize images and capture a wide range of features. This involves transforming a person's face over time. One of the first techniques that produced high-quality results was face aging with IPCGANs. This technique utilizes IPCGANs for face aging. IPCGANs consist of three modules: a CGAN module, a character-safeguard module, and an age classifier. All parts of IPCGANs are differentiable, and the entire architecture can be trained end-to-end.

By utilizing a cutting-edge advancement that employs a novel multi-area picture-to-picture generative adversarial network architecture capable of simulating a seamless bi-directional aging progression within its trained latent space, the model was implemented. The model is trained on the FFHQ dataset, which is annotated for age, orientation, and semantic segmentation. Fixed age classes are used as reference points to approximate continuous age change.

3.3 Image animation module

The Image Animation module takes a source image and a source driving video as inputs and generates a video output that animates the source image based on the driving video. The GUI should provide a reliable and easy-to-use interface for the user to perform tasks specified by the system. The user application, specifically the web application, should offer a dependable and user-friendly interface for users to carry out tasks defined by the system. The application should be easy to use for everyone. The frontend of the application is written in English, and its flow is similar to that of many applications currently in use today. The end user will be able to use the application without any difficulty.

The frontend GUI (see Figure 3) is a web application powered by ReactJS. It provides end users with a user-friendly browsing experience. Users are greeted with a visually appealing page that is free of clutter and offers straightforward options. There is a central drop-down that lists all the possible tasks that the application can perform. They are also provided with a brief description of the task and its objectives. When a task is selected, the corresponding input box is modified to match the required inputs for that task. Once the user has uploaded the required inputs, they can click on the 'Submit' button to execute the task and wait for the output to be displayed on the page.

The process of digitally transforming an individual in a photograph into a short, highly realistic animation that accurately replicates the natural movement of the human face can be achieved using first-order motion models. These are similar to the moving pictures found in newspapers and posters from the Harry Potter world.

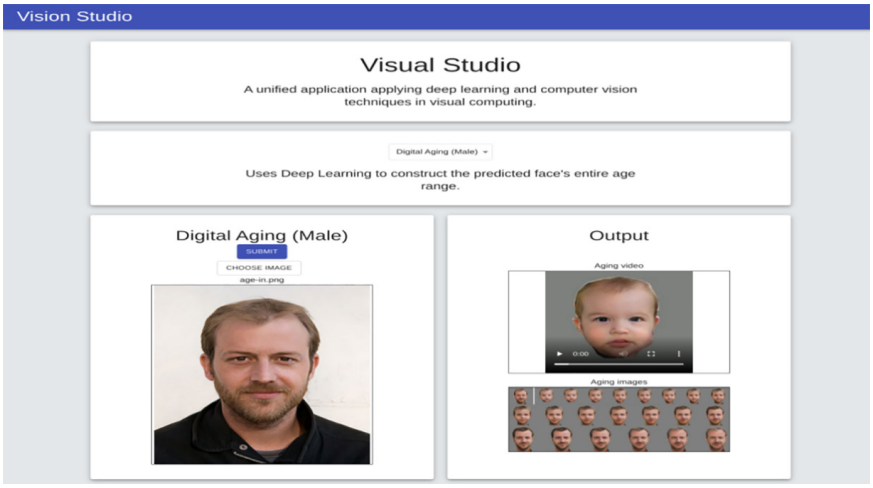

Fig. 3. Graphical user interface

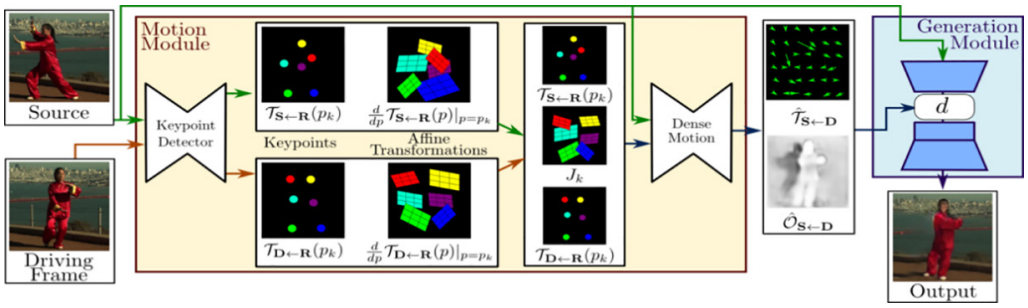

Fig. 4. Explanation of bringing images to life

Picture movement involves creating a video sequence in which an object from a source picture is tracked based on its movement in a reference video. The method examined in [17] to achieve picture liveliness is illustrated in Figure 4. The technique requires a source picture, S, and an edge of a driving video outline, D, as input. The solo key point indicator extracts the initial motion representation, which consists of sparse key points and relative changes with respect to the reference outline R. The dense motion network utilizes this representation to generate a dense optical flow from D to S and an obstacle map. The source picture and the results of the dense motion network are used to produce the target image by the generator. The application, called Vision Studio, provides real-time execution of talks.

3.4 Deepfake detection

It is an interaction that distinguishes deepfakes by utilizing multi-modal identification methods to determine whether an object has been manipulated or artificially created. The most promising technical countermeasures are validation and provenance innovation, along with media literacy measures.

Modern approaches use a combination of two encoder-decoder pairs. The encoder is used to extract latent features of the original image's face, while the decoder reconstructs the face using these latent features. This, of course, requires two encoder-decoder pairs. Both the original and target files are encoded using separate encoders. Once the encoding is complete, the decoders are "swapped," meaning that the source encoder and target decoder are used to generate the final image,

which contains similar inactive elements as the source image. When the training is finished, the decoders are traded and swapped, and a unique encoder of the source picture and a decoder of the target picture are used to reconstruct the target picture using the features of the source picture. The use of such a technique not only preserves the emotions and expressions from the original image but is also much more effective in generating promising and realistic results.

With increased hardware resources, the expectation is that the models will exhibit significantly improved results, allowing for a deeper understanding of how to train them with new data and achieve better optimizations. Each module is separate from the others, allowing for an additional layer of abstraction. By leveraging established methods and optimizing them, the application achieves favorable results in its modules. However, if the hardware resources are limited and the quality of the photos is too high, resulting in blurry grayscale image frames, then our model fails to process the proposed model.

4 RESULTS AND INFERENCES

In general, the EfficientNet models achieve higher accuracy and better efficiency compared to existing CNNs [33], as shown in Figure 5. They reduce parameter size and FLOPS by an order of magnitude. The application utilizes the EfficientNetB7 model as the baseline CNN. For face detection in the model, the BlazeFace library was utilized. This library, developed by Google, is known for its ability to detect faces at incredibly fast speeds. The modules have been implemented using Python, taking into account the fast face detection capabilities of the BlazeFace library. The model was trained end-to-end on the large-scale dataset provided by Kaggle for the Deepfake Detection Challenge. The dataset contains 49 real videos and 49 fake videos, comprising a total of 32,752 frames. The dataset includes a collection of low-quality videos sized 64×64 and another collection of high-quality videos sized 128×128. The dataset is divided into testing and training sets for each quality category, with a 20:80 split.

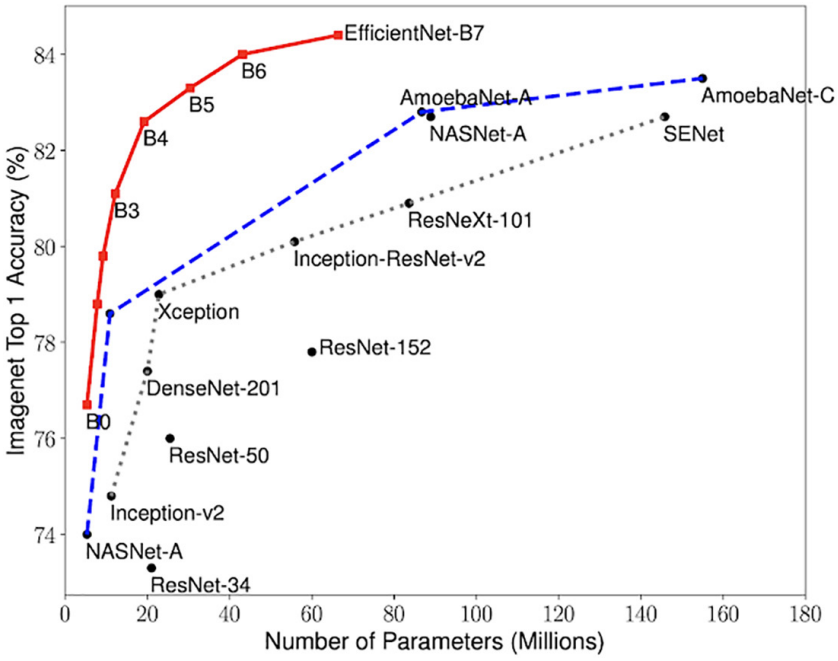

Fig. 5. Graph comparing EfficientNet model accuracy with other state of the art models

Tests were performed separately on each of the existing modules, and the results are explored in the section below. A few tests were done on the modules, and their results are shown below. Other tests were conducted to verify the proper functioning of each module, but they are not presented here. All the tests were conducted on local hardware, and the time taken varied among modules. On significantly better hardware, the modules perform much more efficiently.

Test #1:

Test Input: Deepfake Detection – Video input

Test Expected Output: Score/Probability of video being fake, individual frame scores.

Figure 6 displays a frame capture of the input video used for the deepfake. As observed, there are noticeable alterations in the face, and the probability score is 0.992, suggesting that the model predicts the input video is a deepfake with a 99.2% likelihood. Figure 7 shows several structured frame captures with the likelihood score embedded within the images.

Fig. 6. Deepfake detection test case video input screenshot

Fig. 7. Deepfake detection video output with probability score

Test #2:
Test Input: Image animation – Image and video input
Test Expected Output: Video output

Figure 8 shows the input image that will be animated, while Figure 9 displays a frame capture from the driving video. Figure 10 demonstrates the results of the image animation module test, featuring a few frames captured showcasing the image animated using optical flow from the driving video.

Fig. 8. Image animation input image

Fig. 9. Image animation input driving video (frame capture)

Fig. 10. Frames from output video (image animation)

Test #3:
Test Input: Digital Ageing – Image input
Test Expected Output: Image and sequence video

Figure 11 shows the input image for digital aging, and in Figure 12, it shows the complete range of age images generated from ages 0–70. Additionally, a video sequence is generated, and a frame capture of the video can be seen.

Fig. 11. Digital ageing input

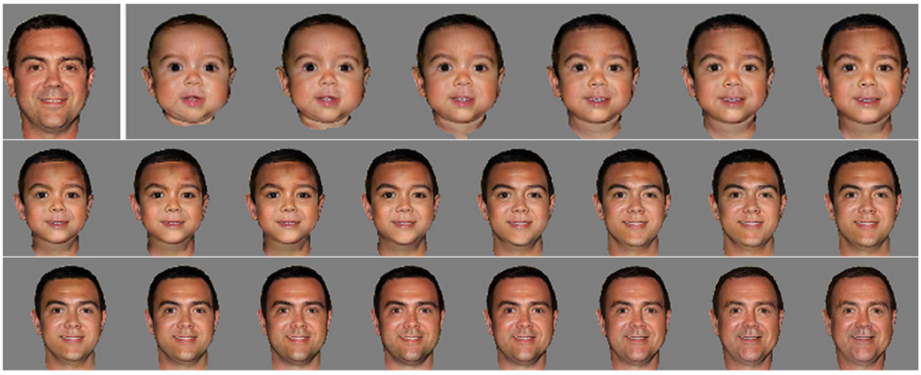

Fig. 12. Digital ageing age range results

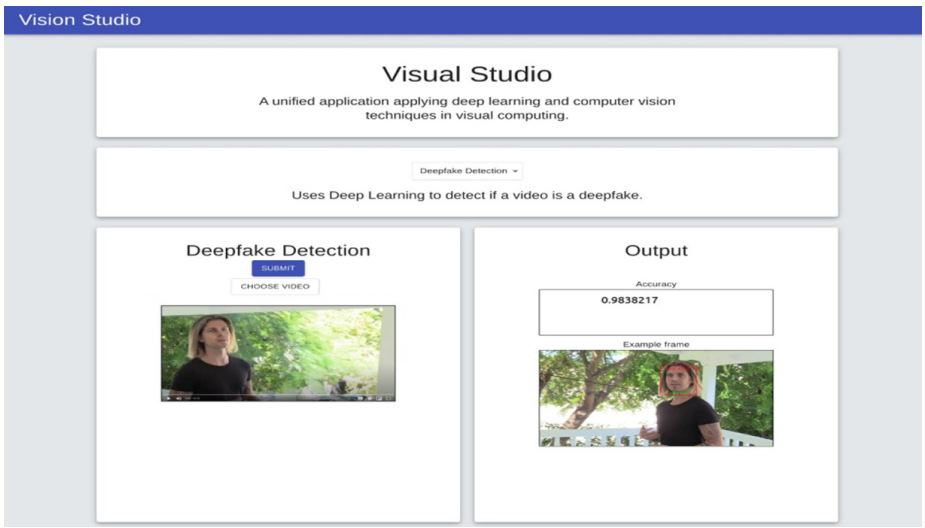

Fig. 13. Frontend of using the deepfake detection module

Figure 13 shows several snapshots from the final GUI, illustrating how the end-user would interact with the frontend. It is completely abstracted from the backend of the application. The frontend results of using the deepfake detection module are as follows: when a deepfake video is given as the input, the output probability is observed to be 0.983. The deepfake detection module indicates that there is a 98.3% probability that the input video is a deepfake. Additionally, it demonstrates the outcome of utilizing the Image Animation module in the frontend. The Image Animation module displays the input image and driving video and provides the end user with the final animated image. This image enhances the input image by utilizing the optical flow from the driving video. Table 1 displays the results of various modules.

Table 1. Results of different modules

Module	Input	Output	Time Taken*	Notes
Digital Ageing (Male)	512 × 512 to 1920 × 1080	256 × 256 to 1920 × 1080 (image, video)	2m12s to 3m45s	Age Range – 0 to 70, Interpolation = 0.5. Any value below this requires more hardware resources.
Digital Ageing (Female)	256 × 256 to 1920 × 1080	256 × 256 to 1920 × 1080 (image, video)	2m27s to 3m47s	Age Range – 0 to 70, Interpolation = 0.5. Any value below this requires more hardware resources.
Image Animation (Face, Pose)	Any resolution video, 512 × 512 image, Video length = 3s to 10s	512 × 512 video	55.812s to 5m33s	Video lengths longer than 10s will require more hardware resources than available locally.
Deepfake Detection	1080 × 720 video (real, fake) Video length = 2s to 30s	Score, Individual frame pictures	8–15s	Highly optimized, the bottleneck is the creation of individual frame pictures for visual clarity.

Note: *The time taken refers to how long it took on local hardware.

5 CONCLUSION

The most important details in this text are the use of deepfakes and realistic media manipulation to disrupt the spread of misinformation and fake news. Vision Studio can help people easily identify deepfakes, thus curbing the spread of misinformation. Deepfakes can also be used for entertainment purposes, such as incorporating imperfections, generating virtual humans for VR games, and creating corporate training videos in the recipient's language. Content creation has multiple vital applications, including virtual reality, videography, gaming, and even retail and advertising. The current development of deep learning and machine learning strategies enables users to transform hours of manual, painstaking content creation work into minutes or even seconds of automated work. Deepfakes can be used to assist in computer-generated content creation, ranging from CGI to game graphics, thereby saving a substantial amount of time. The ability to bring an image to life is mesmerizing, emotional, and highly impactful.

6 CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

7 FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

8 ACKNOWLEDGMENTS

The authors acknowledge the support from MSRIT, REVA University for the facilities provided to carry out the research.

9 REFERENCES

[1] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, Springer Science and Business Media LLC, vol. 2, no. 3, 2021. https://doi.org/10.1007/s42979-021-00592-x

[2] M. Choraś *et al.*, "Advanced machine learning techniques for fake news (online disinformation) detection: A systematic mapping study," *Applied Soft Computing*, Elsevier BV, vol. 101, p. 107050, 2021. https://doi.org/10.1016/j.asoc.2020.107050

[3] Danfeng Xie, Lei Zhang, and Li Bai, "Deep learning in visual computing and signal processing," *Applied Computational Intelligence and Soft Computing*, vol. 2017, no. 1320780, 2017. https://doi.org/10.1155/2017/1320780

[4] N.-A. Perifanis and F. Kitsios, "Investigating the influence of artificial intelligence on business value in the digital era of strategy: A literature review," *Information*, vol. 14, no. 2, p. 85, 2023. https://doi.org/10.3390/info14020085

[5] R. Gil, J. Virgili-Gomà, J.-M. López-Gil, and R. García, "Deepfakes: Evolution and trends," *Soft Computing*, Springer Science and Business Media LLC, vol. 27, no. 16, pp. 11295–11318, 2023. https://doi.org/10.1007/s00500-023-08605-y

[6] F. Matern, C. Riess, and M. Stamminger, "Exploiting visual artifacts to expose deepfakes and face manipulations," in *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, pp. 83–92, 2019. https://doi.org/10.1109/WACVW.2019.00020

[7] Z. Akhtar, "Deepfakes generation and detection: A short survey," *Journal of Imaging*, vol. 9, no. 1. p. 18, 2023. https://doi.org/10.3390/jimaging9010018

[8] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 1–8, 2008. https://doi.org/10.1145/1360612.1360638

[9] Y. Lin, Q. Lin, F. Tang, and S. Wang, "Face replacement with large-pose differences," in *Proceedings of the 20th ACM international conference on Multimedia, ACM*, 2012. https://doi.org/10.1145/2393347.2396426

[10] I. Korshunova, W. Shi, J. Dambre, and L. Theis, "Fast face-swap using convolutional neural networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 3697–3705. https://doi.org/10.1109/ICCV.2017.397

[11] R. Natsume, T. Yatagawa, and S. Morishima, "RSGAN: Face swapping and editing using face and hair representation in latent spaces," *ACM SIGGRAPH 2018 Posters, ACM*, 2018. https://doi.org/10.1145/3230744.3230818

[12] X. Tang, Z. Wang, W. Luo, and S. Gao, "Face aging with identity-preserved conditional generative adversarial networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 7939–7947. https://doi.org/10.1109/CVPR.2018.00828

[13] R. Or-El, S. Sengupta, O. Fried, E. Shechtman, and I. Kemelmacher-Shlizerman, "Lifespan age transformation synthesis," *Computer Vision – ECCV 2020*, Springer International Publishing, pp. 739–755, 2020. https://doi.org/10.1007/978-3-030-58539-6_44

[14] X. Yao, G. Puy, A. Newson, Y. Gousseau, and P. Hellier, "High resolution face age editing," in *25th International Conference on Pattern Recognition (ICPR)*, Milan, Italy, 2021, pp. 8624–8631. https://doi.org/10.1109/ICPR48806.2021.9412383

[15] Y. Alaluf, O. Patashnik, and D. Cohen-Or, "Only a matter of style," *ACM Transactions on Graphics*, vol. 40, no. 4, pp. 1–12, 2021. https://doi.org/10.1145/3450626.3459805

[16] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," in *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018. https://doi.org/10.1109/MSP.2017.2765202

[17] D. P. Kingma and M. Welling, "Auto-encoding variational bayes." *arXiv*, 2013. https://doi.org/10.48550/ARXIV.1312.6114

[18] A. Mallya, T.-C. Wang, K. Sapra, and M.-Y. Liu, "World-consistent video-to-video synthesis," *Computer Vision – ECCV 2020*, Springer International Publishing, pp. 359–378, 2020. https://doi.org/10.1007/978-3-030-58598-3_22

[19] A. Bansal, S. Ma, D. Ramanan, and Y. Sheikh, "Recycle-GAN: Unsupervised video Retargeting," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 119–135, 2018. https://doi.org/10.1007/978-3-030-01228-1_8

[20] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe, "Animating arbitrary objects via deep motion transfer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. https://doi.org/10.1109/CVPR.2019.00248

[21] Y. Xu, F. Xu, Q. Liu, and J. Chen, "Improved first-order motion model of image animation with enhanced dense motion and repair ability," *Applied Sciences*, vol. 13, no. 7, 2023. https://doi.org/10.3390/app13074137

[22] Y. Zhang, L. Zheng, and V. L. L. Thing, "Automated face swapping and its detection," in *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)*, Singapore, 2017, pp. 15–19. https://doi.org/10.1109/SIPROCESS.2017.8124497

[23] P. Korshunov and S. Marcel, "Deepfakes: A new threat to face recognition? Assessment and detection," *arXiv,* preprint arXiv:1812.08685. 2018.

[24] D. Güera, S. Baireddy, P. Bestagini, S. Tubaro, and E. J. Delp, "We need no pixels: Video manipulation detection using stream descriptors," *arXiv*, 2019. https://doi.org/10.48550/ARXIV.1906.08743

[25] P. Korshunov and S. Marcel, "Speaker inconsistency detection in Tampered video," in *26th European Signal Processing Conference (EUSIPCO)*, Rome, Italy, 2018, pp. 2375–2379. https://doi.org/10.23919/EUSIPCO.2018.8553270

[26] Li. Yuezun and Siwei Lyu, "Exposing deepfake videos by detecting face warping artifacts. arXiv 2018," *arXiv,* preprint arXiv:1811.00656 (1811), vol. 2, 2018.

[27] A. A. M. Albazony, H. A. Al-Wzwazy, A. S. Al-Khaleefa, M. A. Alazzawi, M. Almohamadi, and S. E. Alavi, "Deepfake videos detection by using recurrent neural network (RNN)," in *Al-Sadiq International Conference on Communication and Information Technology (AICCIT)*, Al-Muthana, Iraq, 2023, pp. 103–107. https://doi.org/10.1109/AICCIT57614.2023.10217956

[28] D. M. Montserrat, H. Hao, S. K. Yarlagadda, S. Baireddy, R. Shao, J. Horvath, E. Bartusiak, J. Yang, D. Guera, F. Zhu, and E. J. Delp, "Deepfakes detection with automatic face weighting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2020. https://doi.org/10.1109/CVPRW50498.2020.00342

[29] S. J. Pipin, R. Purba, and M. F. Pasha, "Deepfake video detection using spatiotemporal convolutional network and photo response non uniformity," in *IEEE International Conference of Computer Science and Information Technology (ICOSNIKOM)*, Laguboti, North Sumatra, Indonesia, 2022, pp. 1–6. https://doi.org/10.1109/ICOSNIKOM56551.2022.10034890

[30] H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "Multi-task learning for detecting and segmenting manipulated facial images and videos," in *IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Tampa, FL, USA, 2019, pp. 1–8. https://doi.org/10.1109/BTAS46853.2019.9185974

[31] L. Huang and C.-M. Pun, "Audio replay spoof attack detection by joint segment-based linear filter bank feature extraction and attention-enhanced DenseNet-BiLSTM network," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2020, vol. 28, pp. 1813–1825. https://doi.org/10.1109/TASLP.2020.2998870

[32] Z. Wu, R. K. Das, J. Yang, and H. Li, "Light convolutional neural network with feature genuinization for detection of synthetic speech attacks," *arXiv*, 2020. https://doi.org/10.48550/ARXIV.2009.09637

[33] H. V. Ramachandra, Pundalik Chavan, S. Supreeth, H. C. Ramaprasad, K. Chatrapathy, G. Balaraju, S. Rohith, and H. S. Mohan, "Secured wireless network based on a novel dual integrated neural network architecture," *Journal of Electrical and Computer Engineering*, vol. 2023, no. 11, 2023. https://doi.org/10.1155/2023/9390660

10 AUTHORS

Dr. B.J. Sowmya is an Associate Professor in the Department of Artificial Intelligence and Data Science at Ramaiah Institute of Technology. She has 14 years of experience. She has done her M. Tech in year 2013 and P.hd in year 2022. Her area of interests are Deep Learning, Data Analytics, Software Engineering, and Machine Learning. She can be contacted at this email: sowmyabj@msrit.edu.

Dr. Meeradevi is an Associate Professor in the Department of Artificial Intelligence & Machine Learning at Ramaiah Institute of Technology. She has 18 years of working experience. She has done her B.E. in year 2006 & M. Tech. in year 2009 from VTU. Her area of interests are Wireless Sensor Network, Computer Security, Machine Learning, and Computer Networks.

Dr. S. Seema is a Professor in Computer Science and Engineering Department, M S Ramaiah Institute of Technology, Bangalore, India. Her area of research is in Machine Learning and Bioinformatics. Her research interests include Machine Learning, Data Analytics, Virtual Reality and Augmented Reality. She is a member of ACM, and ISTE. She has published over 40 technical papers published in reputed Indian and International Conferences and Journals.

Dr. P. Dayananda is Professor and Head of the Department of Information Technology, Manipal Institute of Technology Bengaluru, Mahe. In previous assignment he was working as Professor and HOD in the Department of Information Science and Engineering at JSSATE, Bengaluru. He Obtained Ph.D degree from VTU and M. Tech degree from RVCE. His focus area is image processing and information retrieval. He has published many papers in national and international journals in the field of image processing and retrieval. He has got a few research grants and consultancy into his account (E-mail: dayananda.p@manipal.edu).

Dr. S. Supreeth is working as an Associate Professor at the School of CSE, REVA University, Bengaluru. He holds Ph.D degree in Computer Science and Engineering from Visvesvaraya Technological University, India. He has been a technology educator for more than ten years. He has published many papers in International journals and conferences in the domain of Cloud, Fog Computing, and Machine Learning.

G. Shruthi holds Ph.D. degree in Computer Science and Engineering from Visvesvaraya Technological University (VTU), Belgaum, and currently she is working as an Assistant Professor in School of Computer Science and Engineering, REVA University, Bengaluru, India. Her areas of interest include Cloud computing, Fog Computing, Machine Learning, and Data Science.

S. Rohith is currently working as an Associate Professor at Department of E&C, NCET, Bengaluru Department of Electronics and Communication Engineering, Nagarjuna College of Engineering and Technology, Bengaluru, Karnataka, India.