CYBR 560 Literature Review

Jeet Kumar

jk32@hood.edu

Fall 2024

The rise of artificial intelligence in the technology industry has produced many marvels in advancement. However, it has also led to the generation of malicious content like the deepfake technology. This method uses deep learning techniques to generate audio and video content which is difficult to distinguish from genuine media. Deepfakes create a unique and consequential threat to the privacy and security of organizations and people alike. This technology creates a unique challenge of identifying, recognizing and discerning real and fake content and has initiated a development of effective detection mechanisms for this task. [1] The process of accurately detecting deepfakes has prompted academics and researchers to investigate various techniques of classification including convolutional neural networks (CNNs) to address this issue.

Deepfake technology is essentially an image classification issue where CNNs have shown considerable success. This technique of extracting and learning from different layers in the pixels in an image makes it a suitable candidate for locating the subtle manipulations within deepfakes. [2]

Among these prominent CNNs is ResNet or Residual Networks, which has introduced the concept of Residual Learning. This facilitates the training of substantially deeper learning networks by establishing a solution for the vanishing gradient. A recent version of ResNet, ResNet RS, balances computational efficiency and performance for image analysis which makes it clear candidate for the task of deepfake detection. [3]

Another CNN that will be used is EfficientNet, which is notable for its scaling of depth, width and resolution. Its main constraint of limiting computational resources is the main metric of performance which contrasts to ResNet. [4] An improved version of EfficientNet, EfficientNetV2, utilizes a more sophisticated scaling system and an enhanced training time for its application to any scenario it is required for. [5]

These are promising candidates when it comes to determining the validity of deepfakes, however, their differences in architecture, computational efficiency and resource consumption may lead to a distinction of performance in relation to their detection efficacy.

Deepfake Detection

Deepfake technology is referring to artificial intelligence that masquerades as real by manipulating original content and manufacturing it to become realistic. By using deep learning techniques, specifically generative adversarial networks (GANs), deepfakes can create realistic audio and video and synthesize new media entirely. [6] Over the past decade, this technology has created a massive trend, which has been driven open-source software and trends in the social media industry, making it easier for amateurs and professional alike to manufacture hyper-realistic content. This has created a threat to the cybersecurity industry as deepfakes establish themselves as a unique tool for any would-be attackers. By generating and synthesizing lifelike content, attacks such as impersonations, social engineering and fraud are all viable options for anyone willing to using to use this technology for malicious purposes. It can also lead to widespread misinformation by creating false news and influence public opinion, especially during important times like elections or political conflicts. As an answer to this evolving challenge, an effective deepfake detection can help identify content that has been forged and protect organizations, people and society from deceitful misconduct and assist with the reduction of the impact of identity theft and misinformation. The development of increasingly robust algorithms and neural network technologies will require advancement as the deepfake technology continues to evolve.

Background of Neural Networks

Neural networks are the foundation of the modern deepfake detection technology we know today. These models are inspired by the learning patterns of the human brain which uses interconnected layers of structured data to create a capability of learning complicated information based on patterns recognized in data. A significant innovation within the neural network field came with the introduction of CNNs. [7] During its inception, this technology was used for rudimentary image analysis by extracting key features within an image, like an edge or a texture, in initial layers and recognizing different models and objects within deeper layers. This level of cognizance makes CNNs suitable for tasks within computer vision, like deepfake detection, where the recognition of subtle manipulations between forged and authentic content is critical. [7]

Background of Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a category of deep learning models that created a foundation for various applications within image analysis. The architecture of a CNN which uses several types of layers to extract features from an input image like textures or patterns. Convolutional and pooling layers are used in tandem to train the filters to highlight areas of importance and reduce the computational requirements. These layers create a feature map which is passed through the layers to establish a link between the translations. This assists with substantiating a final prediction for the image or video being tested. Notable CNNs like AlexNet, ResNet and EfficientNet have been used extensively for the classification of images and achieved remarkable benchmarks. [2] Due to their algorithmic approach to recognizing complex visual patterns, any subtle shifts or transformations in an image or video seldom affects the analysis that is done by the CNN.

EfficientNetV2 is an advanced CNN which enhances the performance of its predecessor, EfficientNet. It uses a model scaling approach to balance network depth, width and resolution which allows it to achieve and maintain a state-of-the-art efficacy. A notable architectural improvement that is utilized is the fused-MBConv layer, which combines the usage of depthwise and pointwise convolutions and transitions this into a single operation which decreases training time. [5] The EfficientNetV2 has been tested rigorously on various datasets where the model showcased an improvement in efficacy and training speed compared to other CNNs like ResNet.

ResNet-RS is an improved version of the original ResNet (Residual Networks) architecture which was created to train deep neural networks at faster pace. One of its notable features is the introduction of residual connections or “skip connections” which allows the network to train properly without suffering from the vanishing gradient issue. [2] By bypassing layers within the network, ResNet is able to avoid a degradation in performance that afflicts other models and also makes it easier for the network to train efficiently. ResNet-RS (Revised Scaling) has added more layers to the model but it compensates for this intensive usage of resources by ensuring the network expands in a more balanced approach. This technique grants the model a more computationally efficient advantage than its predecessor and ensures that an increase in efficacy is not being established at the cost of resources, making it more practical for deepfake detection applications. [3]

Comparative Metrics

These two models are designed with different principles and priorities in terms of complexity, training speed and scalability. While there are multiple studies corroborating their efficacy in real world scenarios, there are few examples of literature that compare the robustness of ResNet RS and EfficientNetV2 to deepfake detection in images or videos. By using a combination of performance metrics, it is possible to gain a holistic understanding of which model is more applicable to the challenges presented by deepfake detection. The total amount of time that is required to train a model can provide an insight into the computational efficiency of the models. The precision and accuracy of the correctly identified deepfakes can be used as a metric to gauge the validity of each model especially if used on the same dataset. Measuring the resource usage of each model will also provide an explanation into the rationality of each model for this particular issue. Using these metrics, a comprehensive analysis can be used as evidence as to the plausibility of these models and technology for this generative issue.

Conclusion

The literature encompassing deepfake detection using convolutional networks (CNNs) like ResNet RS and EfficientNetV2 reflects a significant development in image and video analysis with both architectural families indicating a notable performance in comprehensive classification tasks. While there are gaps in the existing research comparing these two CNNs in the domain of deepfake detection, each excels in a different manner. The focus on efficiency through residual connections that ResNet has and the focus on resource-efficient training that EfficientNetV2 has make them unique tools in the establishing more reliable and scalable deepfake detection systems. There are several characteristics of these two libraries that remain understudied including computational efficiency and detection efficacy especially for environments that are constrained by resources. These intervals of comprehensive research will need to be studied further to provide beneficial insights to the cybersecurity community and industry as a whole and can establish a future safeguard from deepfake technology.

References

|  |  |
| --- | --- |
| [1] | A. A. N. &. A. S. Chawla, "A Survey on Deepfake: Detection Techniques and Applications," *Multimedia Tools and Applications,* vol. 2, no. 81, pp. 2999-3030, 2022. |
| [2] | K. Z. X. R. S. &. S. J. He, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 770-778, 2016. |
| [3] | W. F. X. D. E. D. C. A. S. T.-Y. L. J. S. B. Z. Irwan Bello, "Revisiting ResNets: Improved Training and Scaling Strategies," *Advances in Neural Information Processing Systems,* 2021. |
| [4] | M. Tan, "Efficientnet: Rethinking model scaling for convolutional neural networks," *International Conference on Machine Learning,* pp. 1-11, 2019. |
| [5] | M. a. Q. L. Tan, "Efficientnetv2: Smaller models and faster training," *International conference on machine learning,* pp. pp. 10096-10106, 2021. |
| [6] | J. S. A. M. F. W. Tim Beringer, "Dissecting Convolutional Neural Networks for Runtime and Scalability Prediction".*ICPP '24: Proceedings of the 53rd International Conference on Parallel Processing.* |
| [7] | L. B. Y. B. a. P. H. Yann LeCun, "Gradient Based Learning Applied to Document Recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. pp. 2278-2324, 1998. |