ISSN: 2321-2152

IJMECE

International Journal of modern
electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN2321-2152

www.ijmece .com

Vol 12, Issue.2 April 2024

DETECTION OF DEEPFAKE VIDEOS USING LONG DISTANCE ATTENTION

Muniganti Nikhil Reddy [1], Thudum Srinath [2],
Yanala Jeswanth Reddy [3], Kanukuntla Rithwik[4], Dr. E. John Alex
[1,2,3,4] UG Student, Department of ECE, CMR Institute of Technology, Hyderabad
[5] Professor, Department of ECE, CMR Institute of Technology, Hyderabad

Abstract

With the rapid progress of deepfake techniques in recent years, facial video forgery can generate highly deceptive video contents and bring severe security threats. And detection of such forgery videos is much more urgent and challenging. Mos t existing detection methods treat the problem as a vanilla binary classification problem. In this paper, the problem is treated as a special fine-grained classification problem since the differences between fake and real faces are very subtle. It is observed that most existing face forgery methods left some common artifacts in the spatial domain and time domain, including generative defects in the spatial domain and inter-frame inconsistencies in the time domain. And a spatial-temporal model is proposed which has two components for capturing spatial and temporal forgery traces in global perspective respectively. The two components are designed using a novel long distance attention mechanism. The one component of the spatial domain is used to capture artifacts in a single frame, and the other component of the time domain is used to capture artifacts in consecutive frames. They generate attention maps in the form of patches. The attention method has a broader vision which contributes to better assembling global information and extracting local statistic information. Finally, the attention maps are used to guide the network to focus on pivotal parts of the face, just like other fine-grained classification methods. The experimental results on different public datasets demonstrate that the proposed method achieves the state-ofthe-art performance, and the proposed long distance attention method can effectively capture pivotal parts for face forgery.

Introduction

The deepfake videos are designed to replace the face of one person with another’s. The advancement of generative model s [1]–[4] makes deepfake videos become very realistic. In the meantime, the emergence of some face forgery applications[5]–[7] enables everyone to produce highly deceptive forge d videos. Now, the deepfake videos are flooding the Internet. In the internet era, such technology can be easily used to spread rumors and hatred, which brings great harm to society . Thus the high quality deepfake videos that cannot be distinguished by human eyes directly have aroused interest among researchers. An effective detection method is urgently needed. The general process of generating deepfake videos is shown in Fig. 1. Firstly, the video is divided into frames and the face in each frame is located and cropped. Then, the original face is converted into the target face by using a generative model an d spliced into the corresponding frame. Finally, all frames are serialized to compose the deepfake video. In these processes, two kinds of defects are inevitably introduced. In the process of generating forged faces, the visual artifacts in the spatial domain are introduced by the imperfect generation model. In the process of combining frame sequences into videos, the inconsistencies between frames are caused by the lack of global constraints. Many detection methods are proposed [8]–[10] based on the defects in the spatial domain. Some of the methods take advantage of the defects of face semantics in deepfake videos, because the generative models lack global constraints in th e process of fake face generation, which introduces some abnormal face parts and mismatched details in the face from a global perspective. For example, face parts with abnormal positions [10], asymmetric faces [11], and eyes with different colors [8]. However, it's fragile to rely entirely on these semantics.

Literature survey

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in Advances in Neural Information Processing Systems, vol. 27, Montreal, CANADA, 2014.

We propose a new framework for estimating generative models via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In the space of arbitrary functions G and D, a unique solution exists, with G recovering the training data distribution and D equal to 1 2 everywhere. In the case where G and D are defined by multilayer perceptrons, the entire system can be trained with backpropagation. There is no need for any Markov chains or unrolled approximate inference networks during either training or generation of samples.

D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” 2014.

How can we perform efficient inference and learning in directed probabilistic models, in the presence of continuous latent variables with intractable posterior distributions, and large datasets? We introduce a stochastic variational inference and learning algorithm that scales to large datasets and, under some mild differentiability conditions, even works in the intractable case. Our contributions are two-fold. First, we show that a reparameterization of the variational lower bound yields a lower bound estimator that can be straightforwardly optimized using standard stochastic gradient methods. Second, we show that for i.i.d. datasets with continuous latent variables per datapoint, posterior inference can be made especially efficient by fitting an approximate inference model (also called a recognition model) to the intractable posterior using the proposed lower bound estimator. Theoretical advantages are reflected in experimental results. How can we perform efficient approximate inference and learning with directed probabilistic models whose continuous latent variables and/or parameters have intractable posterior distributions? The variational Bayesian (VB) approach involves the optimization of an approximation to the intractable posterior.

T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive Growing of GANs for Improved Quality, Stability, and Variation,” in International Conference on Learning Representations, Vancouver, Canada, 2018.

We describe a new training methodology for generative adversarial networks. The key idea is to grow both the generator and discriminator progressively: starting from a low resolution, we add new layers that model increasingly fine details as training progresses. This both speeds the training up and greatly stabilizes it, allowing us to produce images of unprecedented quality, e.g., CELEBA images at 10242 . We also propose a simple way to increase the variation in generated images, and achieve a record inception score of 8.80 in unsupervised CIFAR10. Additionally, we describe several implementation details that are important for discouraging unhealthy competition between the generator and discriminator. Finally, we suggest a new metric for evaluating GAN results, both in terms of image quality and variation. As an additional contribution, we construct a higher-quality version of the CELEBA dataset.

Q. Duan and L. Zhang, "Look More Into Occlusion: Realistic Face Frontalization and Recognition With BoostGAN," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 1, pp. 214–228, 2021.

Occlusion in facial photos poses a significant challenge for machine detection and recognition. Consequently, occluded face recognition for camera-captured images has emerged as a prominent and widely discussed topic in computer vision. The present standard face recognition methods have achieved remarkable performance in unoccluded face recognition but performed poorly when directly applied to occluded face datasets. The main reason lies in the absence of identity cues caused by occlusions. Therefore, a direct idea of recovering the occluded areas through an inpainting model has been proposed. However, existing inpainting models based on an encoder-decoder structure are limited in preserving inherent identity information. To solve the problem, we propose ID-Inpainter, an identity-guided face inpainting model, which preserves the identity information to the greatest extent through a more accurate identity sampling strategy and a GAN-like fusing network. We conduct recognition experiments on the occluded face photographs from the LFW, CFP-FP, and AgeDB-30 datasets, and the results indicate that our method achieves state-of-the-art performance in identity-preserving inpainting, and dramatically improves the accuracy of normal recognizers in occluded face recognition. In recent years, occluded face recognition has become a research hotspot in computer vision. Unlike unoccluded faces, occluded faces suffer from incomplete visual components and insufficient identity cues, which lead to degradation in recognition accuracy by normal recognizors [1,2,3,4

"deepfake," http://www.github.com/deepfakes/ Accessed September 18, 2019.

With the rapid progress of deepfake techniques in recent years, facial video forgery can generate highly deceptive video contents and bring severe security threats. And detection of such forgery videos is much more urgent and challenging. Mos t existing detection methods treat the problem as a vanilla binary classification problem. In this paper, the problem is treated as a special fine-grained classification problem since the differences between fake and real faces are very subtle.

EXIXTING SYSTEM

In the past few years, the performance of general image classification tasks has been significantly improved. From the amazing start of Alexnet [31] in Imagenet [32], the method based on deep learning almost dominate the Imagenet competition. However, for fine-grained object recognition [33]–[37], there are still great challenges. The main reason is that the two objects are almost the same from the global and apparent point of visual. Therefore, how to recognize the subtle differences in some key parts is a central theme for fine-grained recognition.

Earlier works [38], [39] leverage human-annotated bounding box of key parts and achieve good results. But the disadvantage is that it needs expensive manual annotation, and the location of manual annotation is not always the best distinguishing area [40], [41], which completely depends on the cognitive level of the annotator.Since the key step of fine-grained classification is focusing on more discriminative local areas [42], many weakly supervised learning methods [23], [40], [43] have been proposed. Most of them use kinds of convolutional attention mechanisms to find the pivotal parts for detection. Fu et al. [43] use a recurrent attention convolutional neural network (RA-CNN) to learn discriminative region attention. Hu et al. [44] propose a channel-wise attention method to model interdependencies between channels. In [40], a multi-attention convolutional neural network is adopted and more fine-grained features can be learned. Hu et al. [23] propose a weakly supervised data augmentation network using attention cropping and attention dropping.

Disadvantages

- The spatial attention model is not designed to capture the artifacts that existed in the spatial domain with a single frame.
- The system not implemented Effectiveness of spatial-temporal model which leads the system less effective.

Proposed System

- The experience of the fine-grained classification field is introduced, and a novel long distance attention mechanism is proposed which can generate guidance by assembling global information.

- It confirms that the attention mechanism with a longer attention span is more effective for assembling global information and highlighting local regions. And in the process of generating attention maps, the non-convolution module is also feasible.

- A spatial-temporal model is proposed to capture the defects in the spatial domain and time domain, according to the characteristics of deepfake videos, the model adopts the long distance attention as the main mechanism to construct a multi-level semantic guidance. The experimental results show that it achieves the state-of-the-art performance.

Advantages

- In the proposed system, the motivation to use long distance attention is given first and then the proposed model is described briefly. As aforementioned, there is no precise global constraint in the deepfake generation model, which always introduces disharmony between local regions in the face forgery from a global perspective.
- In addition to the artifacts that exist in each forgery frame itself, there are also inconsistencies (e.g., unsmooth lip movement) between frame sequences because the deepfake videos are generated frame by frame. To capture these defects, a spatial-temporal model is proposed, which has two components for capturing spatial and temporal defects respectively. Each component has a novel long distance attention mechanism which can be used to assembling the global information to highlight local regions.

Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse Datasets and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Video Detection Type, View Video Detection Type Ratio, Download Predicted Data Sets, View Video Predicted Type Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, Predict Video Detection Type, VIEW YOUR PROFILE.

Decision tree classifiers

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

Gradient boosting

Gradient boosting is a machine learning technique used in regression and classification tasks , among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest.A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

K-Nearest Neighbors (KNN)

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not “learn” until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

Logistic regression Classifiers

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Naïve Bayes

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .

Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).

While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique.

SVM

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point *x* and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space.

CONCLUSION

In this paper, we detect deep fake video from the perspective of fine-grained classification since the difference between fake and real faces is very subtle. According to the generation defects of the deep fake generation model in the spatial domain and the inconsistencies in the time domain, a spatial temporal attention model is designed to make the network focus on the pivotal local regions. And a novel long distance attention mechanism is proposed to capture the global semantic inconsistency in deep fake. In order to better extract the texture information and statistical information of the image, we divide the image into small patches, and recalibrate the importance between them. Extensive experiments have been performed to demonstrate that our method achieves state-of the- art performance, showing that the proposed long distance attention mechanism is capable of generating guidance from a global perspective. Apart from the spatial-temporal model and the long distance attention mechanism, we think a main contribution of this paper is that we confirm not only focusing on pivotal areas is important, but combining global semantics is also critical. This is a noteworthy point, which can be a strategy to improve current models.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” in Advances in Neural Information Processing Systems, vol. 27, Montreal, CANADA, 2014.

[2] D. P. Kingma and M.Welling, “Auto-Encoding Variational Bayes,” 2014.

[3] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive Growing of GANs for Improved Quality, Stability, and Variation,” in International Conference on Learning Representations, Vancouver, Canada, 2018.

[4] Q. Duan and L. Zhang, “Look More Into Occlusion: Realistic Face Frontalization and Recognition With BoostGAN,” IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 1, pp. 214–228, 2021.

[5] “deepfake,” http://www.github.com/deepfakes/ Accessed September 18, 2019.

[6] “fakeapp,” http://www.fakeapp.com/ Accessed February 20, 2020.

[7] “faceswap,” http://www.github.com/MarekKowalski/ Accessed September 30, 2019.

[8] F. Matern, C. Riess, and M. Stamminger, “Exploiting Visual Artifacts to

Expose Deepfakes and Face Manipulations," in IEEE Winter Applications of Computer Vision Workshops, Waikoloa, USA, 2019, pp. 83–92.

[9] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a Compact Facial Video Forgery Detection Network," in IEEE International Workshop on Information Forensics and Security, Hong Kong, China, 2018, pp. 1–7.

[10] X. Yang, Y. Li, H. Qi, and S. Lyu, "Exposing GAN-Synthesized Faces Using Landmark Locations," in Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, Paris, France, 2019, p. 113–118.

[9] Karne, R. K. ., & Sreeja, T. K. . (2023). PMLC- Predictions of Mobility and Transmission in a Lane-Based Cluster VANET Validated on Machine Learning. International Journal on Recent and Innovation Trends in Computing and Communication, 11(5s), 477–483. https://doi.org/10.17762/ijritcc.v11i5s.7109

[10] Radha Krishna Karne and Dr. T. K. Sreeja (2022), A Novel Approach for Dynamic Stable Clustering in VANET Using Deep Learning (LSTM) Model. IJEER 10(4), 1092-1098. DOI: 10.37391/IJEER.100454.

[11] Reddy, Kallem Niranjan, and Pappu Venkata Yasoda Jayasree. "Low Power Strain and Dimension Aware SRAM Cell Design Using a New Tunnel FET and Domino Independent Logic." International Journal of Intelligent Engineering & Systems 11, no. 4 (2018).

[12] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Design of a Dual Doping Less Double Gate Tfet and Its Material Optimization Analysis on a 6t Sram Cells."

[13] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Low power process, voltage, and temperature (PVT) variations aware improved tunnel FET on 6T SRAM cells." Sustainable Computing: Informatics and Systems 21 (2019): 143-153.

[14] Reddy, K. Niranjan, and P. V. Y. Jayasree. "Survey on improvement of PVT aware variations in tunnel FET on SRAM cells." In 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), pp. 703-705. IEEE, 2017