

A SYSTEM, METHOD AND COMPUTER READABLE MEDIUM BASED ON GENERATIVE ADVERSARIAL NETWORK FOR FRAUD DETECTION AND DATA AUGMENTATION IN ELECTRONIC KNOW YOUR CUSTOMER SYSTEM

TECHNICAL FIELD

[0001] The present disclosure relates generally to a system, method and computer readable medium for fraud detection and data augmentation. More particularly, embodiments of the present disclosure are related to a Generative Adversarial Network (GAN) based system, method and computer readable medium for fraud detection and data augmentation in Electronic Know Your Customer (eKYC) applications.

BACKGROUND

[0002] Identity verification, often referred to as the know your customer (KYC) process, is an integral requirement in several fields and industries in the present times. With the advent of technology, KYC processes have also been made available electronically, known as electronic KYC (eKYC). eKYC broadly refers to a paperless KYC procedure wherein the identity of the customer or user is identified and verified based on identification (ID) documents provided electronically. eKYC has gained acceptance and importance in all fields and industries requiring identity verification, to enable effective identification and verification of customers or users based on the information present in the identification documents, such as passports, identity cards or driving licenses. Said fields and industries include without limitation banking and financial services, education sector, insurance sector, staffing and recruiting services, on-demand economies, peer-to-peer (P2P) marketplaces, online communities, gaming sector, telecommunication services, mobility services, health care, real estate and so on.

[0003] Current eKYC solutions broadly require reading the information that is present in the Identification or Identity (ID) document of the customer and determining whether said identification document is a fraudulent one. However, conventional eKYC solutions are unable to accurately perform both of the above-said functions due to lack

of training data. Conventional eKYC solutions typically lack data to train information extraction models, including but without limitation, Optical Character Recognition (OCR) algorithms that extract textual information from ID documents, landmarks detection algorithms that determine if the document has all the correct elements of a specified template, face detection algorithms to detect if an identifying face is properly present in the card, and/ or the like. Further, some solutions may require detection of image data, video data (for example, holograms on cards that appear to change shape as the viewing angle is changed). Due to lack of training data, such conventional eKYC solutions are also unable to effectively determine fraudulent identification documents. The lack of available training data is primarily because of privacy concerns in data protection and privacy regulations and legislations. Additionally, training data available to conventional eKYC solutions are ineffective on account of the mismatch that can happen between the training data and the real data that would actually be provided to the system. For example, training data generally consists of extremely low-quality photos of identity cards, whereas, on actual implementation of the eKYC solutions, high quality photos are provided as input, thereby resulting in issues with the effectiveness of the system once real-world testing begins. Further, conventional eKYC service providers are able to obtain an extremely limited set of data from their customers. As the effectiveness of a system is directly proportional to the amount of training data available, there would undoubtedly be limitations in the system that is developed with the limited available training data.

[0004] Further, conventional solutions handle the information extraction aspect and the fraud aspect as two separate solutions. Accordingly, depending on the approach, the training data for the information extraction aspect may not be suitable for fraud detection aspect thereby limiting the system even further as either the already small dataset needs to be split in two ways, or extra steps have to be taken to obtain extra data specific to the chosen approach which, as explained above, is quite difficult to do.

SUMMARY

[0005] The following summary is provided to facilitate an understanding of some of the innovative features unique to the disclosed embodiment and is not intended to be a full description. A full appreciation of the various aspects of the embodiments

disclosed herein can be gained by taking into consideration the entire specification, claims, drawings, and abstract as a whole.

[0006] The system, method, and computer readable medium disclosed in the present application are advantageous in several ways, including but not limited to streamlining a single Generative Adversarial Network (GAN) pipeline capable of simultaneously augmenting existing data for information extraction and training an improved, state-of-the-art fraud detection model, saving costs, time, manual effort and other resources employed at the stage of pre-processing data, being simpler to set up and manage, augmenting a large pool of synthesized identification or Identity (ID) data capable of training the data/ information extraction model of an electronic know your customer (eKYC) system and development of information extraction algorithms, using the discriminator model of the GAN pipeline for fraud detection with the ability to detect effectively a fraudulent ID document, obtaining an improved electronic know your customer (eKYC) system by making its security model robust against potential fraud IDs that can be generated using artificial intelligence technology and are unmistakably similar to the original IDs.

[0007] The present disclosure relates generally to a system, method and computer readable medium for fraud detection and data augmentation. More particularly, embodiments of the present disclosure are related to a Generative Adversarial Network (GAN) based system, method and computer readable medium for fraud detection and data augmentation in Electronic Know Your Customer (eKYC) applications.

[0008] The present invention discloses a system, method and computer readable medium based on Generative Adversarial Network (GAN) for fraud detection and data augmentation in an Electronic-Know Your Customer (eKYC) system. In the present invention, a generator model and a discriminator model are first sufficiently trained using a GAN. Random noise data is fed to the generator model to generate or synthesize fake images. Initial ground truth ID data of customers and the synthesized fake images are provided to a discriminator model, as a training data. The discriminator provides a classification training result of the input data. One or more losses in the generator model and the discriminator model are determined and the one

or more determined losses are updated in the generator model and the discriminator model using a backpropagation process. The trained generator model is used for data augmentation and training of a data extraction model in the electronic know your customer (eKYC) system. The trained discriminator model is used for enhancing the classification of the security model for fraud detection in the electronic know your customer (eKYC) system.

[0009] The present invention discloses a system, method and computer readable medium providing a single GAN training pipeline that is able to simultaneously tackle two of the largest problems in any identity verification system, such as electronic know your customer (eKYC) system, that is, fraud detection and data augmentation in electronic know your customer (eKYC) applications. Both GAN training and development of information extraction algorithms are done using the single GAN training pipeline disclosed in the present invention. The discriminator model of a GAN of the present invention is able to detect fraudulent IDs even when they are generated through the state-of-the-art means. The generator model of the GAN is used to augment ID data that is difficult to obtain due to privacy and regulatory concerns and hence said augmented ID data can be used to further increase the accuracy of information extraction algorithms such as OCR and landmarks detection.

[0010] In a first aspect of the present disclosure, a system for fraud detection and data augmentation based on Generative Adversarial Network (GAN) in an electronic know your customer (eKYC) system is disclosed. The system may comprise a Generative Adversarial Network (GAN) comprising a generator model input with at least one random noise data for generating at least one synthesized fake image; and a discriminator model input with an input training data comprising the synthesized fake image and at least one ground truth Identity (ID) data of one or more customers for classifying the input training data as a classification training result comprising a real training identity data or a fake training identity data; wherein one or more losses in the generator model and the discriminator model are determined based on the classification training result; wherein a weight associated with each neuron in the generator model and the discriminator model is adjusted using a backpropagation process, based on the determined losses; wherein the generator model and the

discriminator model are trained until the determined loss remains unchanged; wherein the Generative Adversarial Network (GAN) provides a trained generator model after training the generator model using the random noise data and the classification training result, wherein the Generative Adversarial Network (GAN) provides a trained discriminator model after training the discriminator model using the input training data and the classification training result; and the electronic know your customer (eKYC) system used for fraud detection, comprising a data extraction model for extracting at least one text data from an input image used for fraud detection; and a security model for determining whether the input image is a real identity data or a fake identity data; wherein the trained generator model generates a generated data for data augmentation and training of the data extraction model, wherein the trained discriminator model is used as a binary classifier in the security model to determine whether an input data is the real identity data or the fake identity data and provides a classification result.

[0011] According to an embodiment in conjunction to the first aspect of the present disclosure, the generated data is used for data augmentation in the data extraction model.

[0012] According to an embodiment in conjunction to the first aspect of the present disclosure, the ground truth ID data comprises the real identity data and the fake identity data.

[0013] According to an embodiment in conjunction to the first aspect of the present disclosure, the ground truth identity data is a non-inferred data.

[0014] In a second aspect of the present disclosure, a method based on Generative Adversarial Network (GAN) for fraud detection and data augmentation in an electronic know your customer (eKYC) system is disclosed. The method may comprise feeding a generator model with at least one random noise data; generating, by the generator model, at least one synthesized fake image; providing a discriminator model with an input training data comprising the synthesized fake image and at least one ground truth Identity (ID) data of one or more customers; classifying, by the discriminator

model, the input training data as a classification training result comprising a real training identity data or a fake training identity data; determining the losses of the discriminator model and the generator model based on the classification training result; adjusting a weight associated with each neuron in the generator model and the discriminator model using a backpropagation process, based on the determined losses; training the generator model and the discriminator model until the determined loss remains unchanged; obtaining a trained generator model and a trained discriminator model; using a trained generator model for training and data augmentation in a data extraction model of the electronic know your customer (eKYC) system; and using a trained discriminator model for fraud detection in a security model of electronic know your customer (eKYC) system.

[0015] According to an embodiment in conjunction to the second aspect of the present disclosure, training and data augmentation in a data extraction model comprises steps of: generating, by the trained generator, at least one generated data; and training, the data extraction model with the generated data obtained from the trained generator model.

[0016] According to an embodiment in conjunction to the second aspect of the present disclosure, the generated data is used for data augmentation in the data extraction model.

[0017] According to an embodiment in conjunction to the second aspect of the present disclosure, fraud detection in security model of electronic know your customer (eKYC) system comprises steps of: feeding the trained discriminator model with an input data; using the trained discriminator model as a binary classifier in the security model; determining, by the trained discriminator model, whether the input data is a real identity or a fake identity; producing, by the trained discriminator model, a classification result.

[0018] According to an embodiment in conjunction to the second aspect of the present disclosure, the ground truth ID data comprises the real identity and the fake identity.

[0019] According to an embodiment in conjunction to the second aspect of the present disclosure, the ground truth identity data is a non-inferred data.

[0020] In a third aspect of the present disclosure, a computer readable medium storing a program comprising instructions that, when executed by a system, cause the system to perform the method based on Generative Adversarial Network (GAN) for fraud detection and data augmentation in an electronic know your customer (eKYC) system is disclosed.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] The accompanying drawings are included in order to more clearly illustrate the embodiments of the present disclosure. The drawings included herein provide a further understanding of the disclosure, and are incorporated in and constitute a part of the present disclosure. It is appreciable that the drawings illustrate implementations of the disclosure and, together with the description, serve to explain the principles of the disclosure. It is apparent that the drawings are merely some embodiments of the present disclosure, a person having ordinary skill in this field can obtain other figures according to these figures without paying the premise.

[0022] **FIG. 1** is a block diagram of an example unsupervised Generative Adversarial Network (GAN) system **100** used for fraud detection and data augmentation in an Electronic Know Your customer (eKYC) system **118**, in accordance with the disclosed embodiments.

[0023] **FIG. 2** is a block diagram of an example Generative Adversarial Network (GAN) **201** for training a GAN generator model **204** and the GAN discriminator model **210**, in accordance with the disclosed embodiments.

[0024] **FIG. 3** is a block diagram of an example Generative Adversarial Network (GAN) system **300** using a trained generator model **334** and a trained discriminator model **330** in an electronic know your customer (eKYC) system **318**, in accordance

with the disclosed embodiments.

[0025] FIG. 4 is a flow diagram that illustrates an example method **400** for fraud detection and data augmentation, in accordance with the disclosed embodiments.

[0026] FIG. 5 is a flow diagram that illustrates an example method **500** used for training a data/ information extraction model using a trained GAN generator model, in accordance with the disclosed embodiments.

[0027] FIG. 6 is a flow diagram that illustrates an example method **600** of using a discriminator model trained using a GAN in an electronic know your customer (eKYC) system, in accordance with the disclosed embodiments.

[0028] The summary above, as well as the following detailed description of illustrative embodiments, is better understood when read in conjunction with the appended drawings. For the purpose of illustrating the present disclosure, exemplary constructions of the disclosure are shown in the drawings. However, the disclosure is not limited to specific methods and instrumentalities disclosed herein. Moreover, those in the art will understand that the drawings are not to scale. Wherever possible, like elements have been indicated by identical numbers.

DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

[0029] The particular configurations discussed in the following description are non-limiting examples that can be varied and are cited merely to illustrate at least one embodiment and are not intended to limit the scope thereof.

[0030] Generative Adversarial Networks, or GANs are deep-learning based generative models that use two neural networks, pitting one against the other in order to generate new, synthetic instances of data that can pass for real data. Such architectures are widely used in image generation, video generation, and voice generation. One neural network, called generator model, creates data instances, while the other neural network, called discriminator model, evaluates them for authenticity

i.e. the discriminator determines whether each instance of data that it reviews belongs to the actual training dataset or not. The two models, the generator and discriminator, are trained together. The generator generates synthesized or fake data, which along with initial ground truth ID data is provided to the discriminator. The discriminator then classifies the data as real or fake. During training, the discriminator is updated to classify the real and synthesized data as well as possible, and the generator is updated to fool the discriminator as much as possible. Both the generator and discriminator are updated based on the classification results of the discriminator using backpropagation processes or techniques.

[0031] The present invention provides a fake ID generator that would augment the existing dataset to develop the information extraction algorithms, and a state-of-the-art fake ID detector or discriminator for an effective fraud detector in electronic know your customer (eKYC) system.

[0032] **FIG. 1** is a block diagram of an example unsupervised Generative Adversarial Network (GAN) system **100** used for fraud detection and data augmentation in an electronic know your customer (eKYC) system **118**, in accordance with the disclosed embodiments. The system **100** comprises a GAN **101** and the electronic know your customer (eKYC) system **118**. The GAN **101** comprises a GAN generator model **104** and a GAN discriminator model **110**. The electronic know your customer (eKYC) system **118** comprises a security model **114** and an information/ data extraction training model **116**. The electronic know your customer (eKYC) system **118** broadly refers to any electronic identity verification system requiring confirmation and verification of a customer's identity based on provided ID information. The GAN generator model **104** is input with at least one random noise data **102** to generate or synthesize one or more synthesized fake images **106**. The random noise data **102** chosen for the generator model **104** may be a random noise sampling from a random distribution.

[0033] In some embodiments of the present invention, the random noise data **102** may include the initial ground truth ID data, which is the initial set of data that is provided by the client that is directly verified and is labeled correctly as real or fake

data, may contain images of IDs that have varied backgrounds, images of IDs that are taken in multiple lighting solutions, images of IDs that are actual fake documents that are known to have been earlier identified as fraudulent and/ or the like. The generator model **104** generates synthesized fake images **106** by up sampling the sample of random noise data **102** into synthesized fake images **106** that are meaningful images. In an embodiment of the present invention, the generator model **104** is able to create synthesized fake images **106** through varying random noise data **102** as inputted to the generator model **104** from different areas of distribution.

[0034] In some embodiments of the present invention, the generator model **104** generates realistic synthesized fake images **106**. In some embodiments of the present invention, the synthesized fake images **106** generated by the generator model **104** varies in terms of backgrounds, such as texture, color and content, lighting, as real-world input images of customers may not be perfectly lit or have flat background images. In some embodiments of the present invention wherein data from IDs that are actual fake documents that are known to have been earlier identified as fraudulent are inputted to the generator model **104**, said random noise data **102** informs the generator model **104** to synthesized fake images **106** that is in line with such known fraudulent ID data. The GAN discriminator model **110** is input with an input training data comprising one or more synthesized fake images **106** obtained from the GAN generator model **104** and the real image or one or more initial ground truth Identity (ID) data **108** of one or more customers. The initial ground truth ID data **108** is provided by the client as verified, that is either real or fake, but not inferred.

[0035] The synthesized fake images **106** and the initial ground truth ID data **108** are together referred as an input training data **106, 108** for the discriminator model **110**. The discriminator model **110** determines whether the input training data **106, 108** is a real training ID or a fake training ID and provides a classification training result **112**. On account of the varied synthesized fake images **106** generated by the generator model **104**, the discriminator model **110** is capable of efficiently and accurately determining real or fake IDs across a wider sample variation. In some embodiments where the generator model **104** outputs synthesized fake images **106** that is in line with known fraudulent ID data, the discriminator model **110** is further strengthened to

effectively operate and identify real world known fraudulent ID data. In some embodiments, the input training data may comprise multiple images of the same ID so as to train the discriminator model **110** to identify wear and tear in an ID document. In some embodiments of the present invention, the discriminator model **110** is a convolutional neural network (CNN) that has a binary classifier **111** that labels the input data as real ID data or fake ID data. Convolutional neural network is a class of deep neural networks that is most commonly applied for analysing visual data. Accordingly, in an embodiment of the present invention, the discriminator model **110** has a binary classifier **111** to label the input training data as real or fake ID data based on the classification training result **112**. The losses or inaccuracies of the discriminator model **110** and the generator model **104**, which are indicative of the errors in output, are determined, and the weights for the two models are updated, respectively, using a feedback loop or backpropagation process **113**. Backpropagation process **113** refers to a backward propagation of error in a learning of neural networks using gradient descent. In backpropagation process **113**, gradient descent is used to iteratively find a set of parameters or weights for the models that minimises the loss or error. Gradient descent involves using the derivative of a function to adjust the value of a parameter in an attempt to bring it to convergence. Accordingly, by way of backpropagation process **113**, the losses or errors adjust the weights of the generator model **104** so that the generator model **104** can generate more and more realistic synthesized fake images **106** that are closer to the ground truth set of real images as it is trained further. Further, by way of backpropagation process **113**, the losses or errors will adjust the weights of the discriminator model **110** so that said model **110** is able to more effectively verify the authenticity of input training data **106**, **108**. In a context of a neural network, back-propagation process **113** would happen for the weights associated with each neuron in the network, and each iteration of training the neural network will adjust these weights towards minimizing the loss or error of the network. In an embodiment of the present invention, the generator model **104** and the discriminator model **110** are trained until the losses are minimized through backpropagation process **113** (feedback loop).

[0036] Further, as disclosed in **FIG. 1**, once the discriminator model **110** and the generator model **104** are sufficiently trained, the generator model **104** is able to

generate more realistic images that are closer to the ground truth real images, while the discriminator model **110** is able to more accurately and effectively determine whether an input image or data is real or fake.

[0037] The generator model **104** after sufficient training is used to generate synthesized fake images **106** that can augment the data available for the information/data extraction model **116**. The generator model **104** after training is capable of augmenting synthesized images which can be used for training the data/ information extraction model **116** in an electronic know your customer (eKYC) system **118**. Accordingly, in an embodiment of the present invention, the generator model **104** after sufficient training generates synthesized fake images **106** that can augment the data available for the data/ information extraction model **116**. The synthesized fake images **106** that are generated by the generator model **104** after sufficient training, may augment any existing data that the information/data extraction training models may be using at a time. Accordingly, by increasing the data pool of IDs that the information/data extraction algorithms can be trained on, the efficiency of the extraction model is increased phenomenally. Thereby, the data/ information extraction model **116** is capable of extracting at least one text data from an input image used for fraud detection.

[0038] Further, the discriminator model **110** after sufficient training may be used as a binary classifier in the security model **114** of the electronic know your customer (eKYC) system **118** to determine whether the input image **115** of an ID is real or fake, with a high degree of accuracy, as it has been trained with highly realistic fake IDs created by the generator model **104**. Accordingly, by training the discriminator model **110** on realistic synthesized fake images **106** generated by the generator model **104**, the efficiency of the security model **114** is increased phenomenally. The discriminator model **110** after sufficient training can efficiently and correctly determine whether the ID used for electronic know your customer (eKYC) verification is real or fake. Thus, the present invention provides efficient verification of ID for fraud detection in electronic know your customer (eKYC) systems.

[0039] As evident from the above, the present invention provides the advantage of

providing a generator model capable of generating large sets of fake ID data and images that are realistic enough to develop the information extraction algorithms such as OCR systems to read the textual content of the IDs. Further, the present invention also provides a discriminator model that can accurately and robustly detect fraudulent ID documents across a wide span of inputs. Further, by way of backpropagation process or feedback loops in the system, both models are able to improve their efficiency based on the classification determined by the discriminator model in respect of an input image synthesized by the generator model. The present invention thereby solves the problem of limited available training data for electronic know your customer (eKYC) systems in view of privacy concerns and legislations or regulations. A person skilled in the art would recognize that training of the information extraction model of an electronic know your customer (eKYC) system involves extracting text data from an image, also known as Optical Character Recognition (OCR), and to effectively develop an OCR system, a large amount of data is required, especially when it is being developed from the ground up. Hence, augmenting the data pool with realistic images created by the generator model of the present invention greatly enhances the capabilities of the OCR engine or system, as well as saving resources such as time and effort for the client who need not painstakingly source any new data that is already rather difficult to obtain.

[0040] FIG. 2 is a block diagram of an example Generative Adversarial Network (GAN) **201** for training a GAN generator model **204** and the GAN discriminator model **210**, in accordance with the disclosed embodiments. The GAN **201** comprises the GAN generator model **204** and the GAN discriminator model **210**. The generator model **204** is input with at least one random noise data **202** for generating at least one synthesized fake image **206**. The discriminator model **210** is input with an input training data comprising one or more synthesized fake images **206** and one or more initial ground truth Identity (ID) data **208** of customers. The discriminator model **210** is used for classifying the input training data as a classification training result **212** in binary terms, being a real training ID or a fake training ID. The losses or inaccuracies of the discriminator model **210** and the generator model **204**, which are indicative of the errors in output, are determined, and the weights for the two models are updated, respectively, using a feedback loop or backpropagation process **213**.

[0041] The synthesized fake images **206** and the initial ground truth ID data **208** are together referred as an input training data **206, 208** for the discriminator model **210**. The GAN **201** provides a trained generator model after training the generator model **204** using the random noise data **202** and the classification training result **212** and also provides the trained discriminator model after training the discriminator model **210** using the input training data **206, 208** and the classification training result **212**. The generator model **204** and the discriminator model **210** in **FIG. 2** refer to the respective models before or during training. **FIG. 3**, as described below, discloses a trained generator model **334** and a trained discriminator model **330** which in some embodiments, refer to the generator model **204** and the discriminator model **210** obtained after sufficient training.

[0042] The discriminator model **210** also referred to as the discriminator, is a Convolutional Neural Network (CNN) that has a binary classifier that labels an input image as real or fake. The generator model **204** also referred to as generator can be understood to be an inverse of the CNN, that takes in noise input to generate a synthesized image. The distribution of noise has little effect on the final results, hence any kind of random distribution can be used to sample from. Through varying the random noise in the input by sampling from different areas of the distribution, the generator is able to create a wide variety of data.

[0043] The present invention uses the initial ground truth ID data **208** which are IDs of documents that are directly verified and labelled correctly, as real or fake data. In an embodiment of the present invention, the initial ground truth ID data **208** is used in the training of the discriminator model **210** and contains features, such as, images of IDs having varied backgrounds, images of IDs that are taken in multiple lighting situations and images of IDs that are actual fake documents that fraudsters have tried to use in electronic know your customer (eKYC) systems. In some embodiments of the present invention, the synthesized fake images **206** generated by the generator model **204** varies in terms of backgrounds, such as texture, color and content, lighting, as real-world input images of customers may not be perfectly lit or have flat background images. The features of images of IDs having varied backgrounds and

images of IDs that are taken in multiple lighting situations help to expand the generator model **204** to be able to output a more varied set of data. This would also allow the discriminator model **210** to be much more effective at determining real or fake IDs across a wider sample variation. In some embodiments of the present invention, the features of images of fake IDs used by fraudsters would allow actual, ground truth fake ID data that informs the generator model **204** to output fake ID data that is in line with what has been actually used in real life, which further increases the efficiency of the discriminator model **210** in real world situations.

[0044] The initial ground truth ID data **208** is required as a starting pool of data to begin the training process for the base model of the discriminator to determine whether the input image of an ID is real or fake, which can then be incorporated into the GAN pipeline or GAN **201** to train the generator model **204**. However, by way of the present invention, the requirement of the data required to be input by the client is reduced markedly as the generator model is able to augment the existing data available to develop information extraction algorithms.

[0045] Further, the initial ground truth ID data **208** is primarily used to train the discriminator model **210** to tell apart real or fake IDs. The input that is typically chosen for the generator model **204** is a random noise data **202** or the random noise sampling from a random distribution. The training of the generator model **204** is informed by the classification of real/fake by the discriminator model **210**, so the losses/errors will adjust the weights of the generator model **204** so that the generator model **204** can generate more and more realistic synthesized fake images **206** that are closer to the ground truth set of real images as it is trained further.

[0046] **FIG. 3** is a block diagram of an example Generative Adversarial Network (GAN) system **300** using a trained generator model **334** and a trained discriminator model **330** in an electronic know your customer (eKYC) system **318**, in accordance with the disclosed embodiments. The trained generator model **334**, being sufficiently trained, generates one or more generated data **320** that are more realistic images that are closer to ground truth real images. A data extraction model **316** is trained using the generated data **320** obtained from the trained generator model **334**. The data

extraction model **316** is also referred to as the information extraction system or pipeline or data extraction system or pipeline or data extraction training model or system or pipeline.

[0047] In an embodiment of the present invention, the generator model **104** and the discriminator model **110** of **FIG. 1** are trained until the losses are minimized through backpropagation process (feedback loop), and the same may be referred to as the trained generator mode **334** and the trained discriminator model **310**.

[0048] The trained generator model **334** is used to generate new fake data or generated data **320** that can augment the data available for the information/ data extraction training pipelines or model **316**, while the trained discriminator model **330** is incorporated into a security pipeline or model **314**. The data that is generated by the trained generator model **334**, that is, the generated data **320** augments any existing data that the information/ data extraction training pipelines or model **316** are currently using. Basically, it increases the data pool of IDs that the information/data extraction model **316** can be trained on.

[0049] In some embodiments of the present invention, the generator model **204** of **FIG. 2** may be sufficiently trained and may be suitable for data augmentation in information/ data extraction training model **316**. The security pipeline also referred to as security model **314** is a component of the electronic know your customer (eKYC) system **118** that is responsible for determining whether or not the input ID is a fraudulent one. A typical security model **314** can contain several algorithms, for example, an algorithm to detect landmarks in an ID of logos, images or words, that are supposed to be present in a specific ID type and/ or an algorithm to detect holograms in an ID. Adding the sufficiently trained discriminator model **330** into the entire pipeline or entire model makes the entire security model **314** more robust, as now it is protected against potential fraudulent IDs that can be created using Artificial Intelligent (AI) methods that can be extremely similar to an original ID.

[0050] In some embodiments of the present invention, the data extraction model and security model may not be a part of the eKYC system. In such a case, the eKYC

service provider may use a data extraction model or a security model or both from other or external sources or providers including, but not limited to, external agencies facilitating eKYC services, data verification service providers, banks, financial institutions, third party intermediaries and/or other service providers or agencies.

[0051] FIG. 4 is a flow diagram that illustrates an example method **400** for fraud detection and data augmentation, in accordance with the disclosed embodiments. As disclosed in Fig. 4, the example method **400** can be implemented by feeding in a generator model, at least one random noise data, as at block **402**. The input chosen for the generator may be a random noise sampling from a random distribution, as typical in a Generative Adversarial Network (GAN). In some embodiments of the present invention, the initial ground truth ID data, which is the initial set of data that is provided by the client that is directly verified and is labeled correctly as real or fake data, may contain images of IDs that have varied backgrounds, images of IDs that are taken in multiple lighting solutions, images of IDs that are actual fake documents that are known to have been earlier identified as fraudulent and/ or the like. Thereafter, as at block **404**, the generator model generates one or more synthesized images. The synthesized images are generated by the generator model by up sampling the sample of random noise into meaningful images. In some embodiments of the present invention, through varying random noise in the input by sampling from different areas of distribution, the generator is able to create a wide variety of data. In some embodiments of the present invention, the generator model generates realistic synthesized identification (ID) documents or ID images. By inputting a varied initial set of data, the generator model is able to output a more varied set of data, with a wide range of backgrounds (in terms of texture, color and content) as well as lighting, which mimics real world usage which may not be perfectly lit, flat background images in some input training data. In some embodiments wherein images of IDs that are actual fake documents that are known to have been earlier identified as fraudulent are inputted to the generator model, said data informs the generator model to output fake ID data that is in line with such known fraudulent ID data.

[0052] Thereafter, as at block **406**, a discriminator model is provided with an input training data comprising the synthesized fake image obtained from the generator model and one or more initial ground truth ID data of a customer. The initial ground

truth ID data is data that is provided by the client as verified, that is observed to be real or fake, but not inferred. On account of the varied set of data generated by the generator model, the discriminator model is much more effective and efficient at determining real or fake IDs across a wider sample variation. In some embodiments where the generator model outputs fake ID data that is in line with known fraudulent ID data, the discriminator is further strengthened to effectively operate and identify known fraudulent ID data. In some embodiments, the input training data may comprise multiple data of the same ID so as to train the discriminator to identify wear and tear in an ID document. The discriminator model classifies the input training data and provides a classification training result, as at block **408**. In an embodiment of the present invention, the discriminator model is a convolutional neural network (CNN) that has a binary classifier that labels the input training data as real or fake. Accordingly, in an embodiment of the present invention, the classification training result can be a real training ID or a fake training ID. Thereafter, at block **410**, the losses or inaccuracies of the discriminator model and the generator model are determined. A weight associated with each neuron in the generator model and the discriminator model is adjusted using a backpropagation process, based on the determined losses, as at block **412**. Backpropagation refers to a backward propagation of error in a learning of neural networks using gradient descent. In backpropagation, gradient descent is used to iteratively find a set of parameters or weights for the models that minimises the loss or error. Gradient descent involves using the derivative of a function to adjust the value of a parameter in an attempt to bring it to convergence. Accordingly, by way of backpropagation, the losses or errors adjust the weights of the generator model so that the generator model can generate more and more realistic images that are closer to the ground truth set of real images as it is trained further. Further, by way of backpropagation, the losses or errors will adjust the weights of the discriminator model so that said model is able to more effectively verify the authenticity of an input image. The generator model and the discriminator model are trained until the losses are minimized or unchanged through backpropagation process (feedback loop), as at step **414**. Thereafter, the method **400** involves obtaining a trained generator model and the trained discriminator model with minimum or unchanged loss, as at block **416**. As at block **418**, the trained generator model can be used for training and data augmentation in a data extraction model of an electronic know your customer (eKYC) system. As at step **420**, the trained discriminator model can be used for fraud detection

in a security model of electronic know your customer (eKYC) system.

[0053] **FIG. 5** is a flow diagram that illustrates an example method **500** used for training a data/ information extraction model using a trained GAN generator model, in accordance with the disclosed embodiments. As at block **502**, the trained generator model is capable of augmenting synthesized images which can be used for training the data/ information extraction model in an electronic know your customer (eKYC) system. The trained generator model is used to generate new fake (synthesized) data that can augment the data available for the data/ information extraction models. The synthesized data that is generated by the trained generator model may augment any existing data that the information/data extraction training pipelines are currently using. As at block **504**, the data extraction model of the electronic know your customer (eKYC) system is trained with the generated data obtained from the trained generator model. Since the generator model is sufficiently trained, that is, the weights of the generator model are adjusted based on the losses or errors, as described in the method **400** of **FIG. 4**, the generator model is able to generate realistic images that are closer to the ground truth set of real images. Accordingly, by increasing the data pool of IDs that the information/ data extraction algorithms can be trained on, the efficiency of the extraction model is increased phenomenally.

[0054] Some examples of information or data extraction models include without limitation, Optical Character Recognition (OCR) algorithms that extract textual information from identification documents, landmarks detection algorithms that determine whether the card or ID has all the correct elements of a specified template, face detection algorithms to detect if an identifying face is properly present in the card, and so on. In some embodiments of the present invention, the trained generator model may be used to efficiently train any of the above data extraction models or systems. Augmenting the data pool with realistic images created by the generator would greatly enhance the capabilities of the models or systems, for example the OCR engine, as well as save resources, such as time and effort for the user who will not need to painstakingly source for new data that is already rather difficult to obtain.

[0055] **FIG. 6** is a flow diagram that illustrates an example method **600** of using a

discriminator model trained using a GAN in an electronic know your customer (eKYC) system, in accordance with the disclosed embodiments. In an embodiment of the present invention, the trained discriminator model is used as a binary classifier in the security model of the electronic know your customer (eKYC) system. The trained discriminator model is input with an input image or data, as at block **602**. The input image or data can be the image data of ID of one or more customers used at the time of offline examination or verification of signatures in the electronic know your customer (eKYC) system. As at block **604**, the trained discriminator model is used as a binary classifier in a security model of the electronic know your customer (eKYC) system. As at block **606**, the trained discriminator model classifies the input ID image or ID data as a real ID or a fake ID. As at block **608**, the security model of the electronic know your customer (eKYC) system provides the classification result of the discriminator model, that is, real ID or fake ID to the electronic know your customer (eKYC) system. Since the discriminator model is sufficiently trained using a GAN and the weights of the discriminator model are adjusted based on the losses or errors, as described in the method **400** of **FIG. 4**, the discriminator model is able to accurately and robustly detect fraudulent ID data across a wide span of inputs. Accordingly, by training the discriminator model on realistic images generated by the generator model, the efficiency of the security model is increased phenomenally. The sufficiently trained discriminator can efficiently and correctly determine whether the ID used for electronic know your customer (eKYC) verification is real or fake. Thus, the present invention provides efficient verification of ID for fraud detection in electronic know your customer (eKYC) systems.

[0056] The present invention also discloses a computer readable medium storing a program comprising instructions that, when executed by a system, cause the system to perform the method based on Generative Adversarial Network (GAN) for fraud detection and data augmentation in an electronic know your customer (eKYC) system. In one embodiment of the invention, one or more computing devices comprising the computer readable medium is used to perform the methods **400**, **500** and **600** disclosed in **FIG. 4**, **FIG. 5** and **FIG. 6**, respectively.

[0057] The present invention discloses a system, method and computer readable

medium providing a single GAN training pipeline that is able to simultaneously tackle two of the largest problems in any identity verification system, such as electronic know your customer (eKYC) system, that is, fraud detection and data augmentation in electronic know your customer (eKYC) applications. Both GAN training and development of information extraction algorithms are done using the single GAN training pipeline disclosed in the present invention. The discriminator model of a GAN of the present invention is able to detect fraudulent IDs even when they are generated through the state-of-the-art means. The generator model of the GAN is used to augment ID data that is difficult to obtain due to privacy and regulatory concerns and hence said augmented ID data can be used to further increase the accuracy of information extraction algorithms such as OCR and landmarks detection.

[0058] The system, method and computer readable medium of the present invention can create data that can be augmented in developing data extraction algorithms such as OCR, landmarks detection. Further, said system, method and computer readable medium can also create a neural network that can accurately determine whether an ID is fraudulent. Accordingly, one advantage of the present invention is that the invention overcomes a conventional need of two separate projects to develop the extraction algorithms and security model, thereby saving costs, time, manual effort and other resources employed at the stage of pre-processing data. Furthermore, the system, method and computer readable medium of the present invention for determining fake IDs, especially as the electronic know your customer (eKYC) market grows and the fraudsters become more sophisticated, overcomes several disadvantages of the conventional techniques such as traditional image processing techniques.

[0059] It will be appreciated that variations of the above disclosed and other features and functions, or alternatives thereof, may be desirably combined into many other different systems or applications. Also, various presently unforeseen or unanticipated alternatives, modifications, variations or improvements therein may be subsequently made by those skilled in the art which are also intended to be encompassed by the following claims.

[0060] Although embodiments of the current disclosure have been described comprehensively in considerable detail to cover the possible aspects, those skilled in the art would recognize that other versions of the disclosure are also possible.

CLAIMS

What is claimed is:

1. A system based on Generative Adversarial Network (GAN) for fraud detection and data augmentation in an Electronic Know Your customer (eKYC) system, the system comprising:

a Generative Adversarial Network (GAN) comprising:

a generator model input with at least one random noise data for generating at least one synthesized fake image; and

a discriminator model input with an input training data comprising the synthesized fake image and at least one ground truth identity data of one or more customers, for classifying the input training data as a classification training result comprising a real training identity data or a fake training identity data;

wherein one or more losses in the generator model and the discriminator model are determined based on the classification training result;

wherein a weight associated with each neuron in the generator model and the discriminator model is adjusted using a backpropagation process, based on the determined losses;

wherein the generator model and the discriminator model are trained until the determined loss remains unchanged;

wherein the Generative Adversarial Network (GAN) provides a trained generator model after training the generator model using the random noise data and the classification training result,

wherein the Generative Adversarial Network (GAN) provides a trained discriminator model after training the discriminator model using the input training data and the classification training result; and

the electronic know your customer (eKYC) system used for fraud detection, comprising:

a data extraction model for extracting at least one text data from an input image used for fraud detection; and

a security model for determining whether the input image is a real identity data or a fake identity data;

wherein the trained generator model generates a generated data for data augmentation and training of the data extraction model,

wherein the trained discriminator model is used as a binary classifier in the security model to determine whether an input data is the real identity data or the fake identity data and provides a classification result.

2. The system of claim 1, wherein the generated data is used for data augmentation in the data extraction model.

3. The system of claim 1, wherein the ground truth identity data comprises the real identity data and the fake identity data.

4. The system of claim 1, wherein the ground truth identity data is a non-inferred data.

5. A method based on Generative Adversarial Network (GAN) for fraud detection and data augmentation in an electronic know your customer (eKYC) system, the method comprising:

feeding a generator model with at least one random noise data;

generating, by the generator model, at least one synthesized fake image;

providing a discriminator model with an input training data comprising the synthesized fake image and at least one ground truth Identity (ID) data of one or more customers;

classifying, by the discriminator model, the input training data as a classification training result comprising a real training identity data or a fake training identity data;

determining the losses of the discriminator model and the generator model based on the classification training result;

adjusting a weight associated with each neuron in the generator model and the discriminator model using a backpropagation process, based on the determined losses;

training the generator model and the discriminator model until the determined

loss remains unchanged;

obtaining a trained generator model and a trained discriminator model;

using the trained generator model for training and data augmentation in a data extraction model of the electronic know your customer (eKYC) system; and

using the trained discriminator model for fraud detection in a security model of electronic know your customer (eKYC) system.

6. The method of claim **5**, wherein training and data augmentation in the data extraction model comprises steps of:

generating, by the trained generator, at least one generated data; and

training, the data extraction model with the generated data obtained from the trained generator model.

7. The method of claim **6**, wherein the generated data is used for data augmentation in the data extraction model.

8. The method of claim **5**, wherein fraud detection in the security model of electronic know your customer (eKYC) system comprises steps of:

feeding the trained discriminator model with an input data;

using the trained discriminator model as a binary classifier in the security model;

determining, by the trained discriminator model, whether the input data is a real identity data or a fake identity data;

producing, by the trained discriminator model, a classification result.

9. The method of claim **5**, wherein the ground truth identity data comprises the real identity data and the fake identity data.

10. The method of claim **5**, wherein the ground truth identity data is a non-inferred data.

11. A computer readable medium storing a program comprising instructions that, when executed by a system, cause the system to perform the method based on Generative Adversarial Network (GAN) for fraud detection and data augmentation in an electronic know your customer (eKYC) system, of any one of claims **5 - 10**.