Archive of SID.ir

Journal of Modeling in Engineering 21 (2023) 19-28

Semnan University

Journal of Modeling in Engineering

Journal homepage: https://modelling.semnan.ac.ir/

ISSN: 2783-2538



Research Article

Deepfake Image Detection Using a Deep Hybrid Convolutional Neural Network

Fahimeh Bagherzadeh [1], Razieh Rastgoo [2*]

[a] *B.Sc. Student, Faculty of Electrical and Computer Engineering, Semnan University, Semnan, Iran*
[b] *Assistant Professor, Faculty of Electrical and Computer Engineering, Semnan University, Semnan, Iran*

PAPER INFO

Paper history:

Received: 05 August 2023
Revised: 14 October 2023
Accepted: 03 December 2023

Keywords:

Deepfake;
Deep learning; Deep convolutional network;
Accuracy;
Fake images.

ABSTRACT

Recent advances in the field of Artificial Intelligence, particularly in deep learning, have led to significant achievements in various domains. However, in some areas, these advancements have posed threats to individuals' privacy. For instance, one emerging application based on deep learning is Deepfake. Deepfake algorithms can generate synthetic images and videos that humans cannot distinguish from real ones. In this context, it becomes imperative to introduce models and algorithms capable of automatically differentiating between real and fake data. In recent years, numerous studies have been conducted to understand the functioning of Deepfakes, and various deep learning-based methods have been introduced to identify and distinguish videos or images generated by Deepfakes from real ones. To enhance the accuracy of Deepfake detection and concurrently leverage the capabilities of different types of convolutional neural networks, this paper proposes a hybrid model using four convolutional neural networks. Relying on the high capabilities of these networks in extracting effective features from the input image, the proposed model can simultaneously discern the authenticity of the input image through the collaboration of these four models. The presented results on three databases, namely 140k real and fake faces, DFDC faces, and Deepfake and real images, demonstrate accuracies of 99.12%, 96.24%, and 98.80%, respectively. These results indicate an improvement of 2.42%, 9.72%, and 0.55% compared to existing models.

DOI: https://doi.org/ 10.22075/jme.2023.31438.2511

© 2023 Published by Semnan University Press.
This is an open access article under the CC-BY 4.0 license.(https://creativecommons.org/licenses/by/4.0/)

*** Corresponding author.**
E-mail address: rrastgoo@semnan.ac.ir

How to cite this article:
Bagherzadeh, F., & Rastgoo, R. (2023). Deepfake image detection using a deep hybrid convolutional neural network. Journal of Modeling in Engineering, 21(75),19-28. doi: 10.22075/jme.2023.31438.2511

Archive of SID.ir

Archive of SID.ir

مقاله پژوهشی

تشخیص دیپ‌فیک در تصویر با استفاده از مدل ترکیبی مبتنی بر شبکه‌های عصبی کانولوشنی عمیق

فهیمه باقرزاده[۱]، راضیه راستگو[۲،*]

اطلاعات مقاله

دریافت مقاله: ۱۴۰۲/۰۵/۱۴
بازنگری مقاله: ۱۴۰۲/۰۷/۲۲
پذیرش مقاله: ۱۴۰۲/۰۹/۱۲

واژگان کلیدی:
دیپ‌فیک،
یادگیری عمیق،
شبکه کانولوشنی عمیق،
دقت،
تصاویر ساختگی.

چکیده

پیشرفت‌های اخیر در حوزه هوش مصنوعی و به خصوص یادگیری عمیق در بسیاری از زمینه‌ها منجر به کسب نتایج چشمگیری گردیده است. با این حال، در برخی زمینه‌ها نیز این پیشرفت‌ها حریم خصوصی افراد را مورد تهدید قرار داده‌اند. به عنوان نمونه، یکی از الگوریتم‌های کاربردی مبتنی بر یادگیری عمیق که اخیراً ظهور کرده است، دیپ فیک می‌باشد. الگوریتم‌های دیپ‌فیک می‌توانند تصاویر و ویدیوهای جعلی ایجاد کنند که انسان‌ها نمی‌توانند آن‌ها را از نمونه‌های واقعی تشخیص دهند. در این راستا، ارائه مدل‌ها و الگوریتم‌هایی که بتوانند به طور خودکار داده‌های واقعی را از داده‌های جعلی تشخیص دهند، ضروری به نظر می‌رسد. در سال‌های اخیر، مطالعات زیادی برای درک نحوه عملکرد دیپ‌فیک‌ها انجام شده است و روش‌های بسیاری مبتنی بر یادگیری عمیق برای شناسایی ویدیوها یا تصاویر تولید شده توسط دیپ‌فیک و نیز تمایز آنها از تصاویر واقعی معرفی شده است. به منظور بهبود دقت تشخیص دیپ‌فیک و نیز استفاده همزمان از قابلیت-های انواع مختلف شبکه‌های عصبی کانولوشنی، در این مقاله، یک مدل ترکیبی با استفاده از چهار شبکه عصبی کانولوشنی ارائه می‌گردد. با تکیه بر قابلیت‌های بالای این شبکه‌ها در استخراج ویژگی‌های موثر از تصویر ورودی، مدل پیشنهادی قادر به تشخیص همزمان جعلی یا واقعی بودن تصویر ورودی توسط این چهار مدل می‌باشد. نتایج ارائه شده بر روی سه پایگاه داده 140k Real and Fake Faces، DFDC Faces و Deepfake and Real Images به ترتیب برابر با ۹۹.۱۲٪، ۹۶.۲۴٪ و ۹۸.۸۰٪ می‌باشد که نشان‌دهنده میزان ۲.۴۲٪، ۹.۷۲٪ و ۰.۵۵٪ بهبود در نتایج نسبت به مدل‌های موجود می‌باشد.

DOI: https://doi.org/ 10.22075/jme.2023.31438.2511

© 2023 Published by Semnan University Press.
This is an open access article under the CC-BY 4.0 license.(https://creativecommons.org/licenses/by/4.0/)

۱-مقدمه[۱]

هوش مصنوعی در طی سال‌های اخیر توسعه و پیشرفت‌های چشمگیر و قابل‌توجهی داشته و در حوزه‌های مختلف مورد استفاده قرار گرفته است [۱-۱۱]. یادگیری عمیق، به عنوان زیر مجموعه‌ای از هوش مصنوعی، انقلابی را در طی سال های اخیر ایجاد کرده است. علیرغم کاربردهای بسیار مدل‌های یادگیری عمیق در حوزه‌های مختلف [۱۲-۱۳]، این مدل‌ها همیشه برای اهداف مثبت و سازنده استفاده نشده‌اند. به عنوان نمونه، مدل‌های مبتنی بر یادگیری عمیق قادر می‌باشند داده‌های جعلی و مصنوعی از انسان را به صورت واقع‌گرایانه تولید و سنتز نماید [۱۴]. در این راستا، اصطلاح دیپ فیک برای اولین بار در سال ۲۰۱۷ توسط

* پست الکترونیک نویسنده مسئول: rrastgoo@semnan.ac.ir
۱. دانشجوی کارشناسی، دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان
۲. استادیار، دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان

استناد به این مقاله:
باقرزاده، فهیمه، & راستگو، راضیه. (۱۴۰۲). تشخیص دیپ فیک در تصویر با استفاده از مدل ترکیبی مبتنی بر شبکه عصبی کانولوشنی عمیق. مدل سازی در مهندسی، ۲۱(۷۵)، ۱۹-۲۸. doi: 10.22075/jme.2023.31438.2511

Archive of SID.ir

Archive of SID.ir

یکی از کاربران شبکه‌های اجتماعی با نام دیپ‌فیک مطرح گردید. از آن زمان به بعد، دیپ‌فیک به دسته‌ای از داده‌های جعلی و مصنوعی [۱] اطلاق می‌گردد که در آن محتوای جعلی بر اساس محتوای موجود تولید می‌گردد. این محتوا می‌تواند شامل تصویر، ویدیو و سیگنال‌های صوتی باشد. شبکه‌های زایشی عمیق برای تولید محتوای دیپ‌فیک استفاده می‌شوند. داده‌های جعلی کاربردهای مختلفی در جرایم سایبری [۱۳] شامل سرقت هویت، اخبار جعلی، تحریک خشونت، کلاهبرداری مالی، زورگویی سایبری، ویدیوهای فحاشی جعلی افراد مشهور [۱۴] برای باج‌گیری، انتخابات دموکراتیک و بسیاری موارد دیگر می‌باشد. علاوه بر این، دیپ‌فیک برای تهدید سازمان‌ها و افراد نیز استفاده می شود. بر اساس گزارش‌های موجود، در بسیاری از موارد، محتواهای دیپ‌فیک، غیراخلاقی می‌باشند [۱۵]. در این راستا نیازمند ابزار و مدل‌های پیشرفته و دقیق جهت شناسایی جرایم مرتبط با دیپ‌فیک هستیم.

با توجه به بهبود در عملکرد و پیشرفت‌های حاصل‌شده در مدل‌های دیپ‌فیک در طی سال‌های اخیر، نگرانی‌هایی از بابت استفاده نادرست از آنها به وجود آمده است [۱۵]. به عنوان نمونه، این مدل‌ها می‌توانند یک فرد یا افراد را در حال گفتن صحبت‌ها یا انجام اقداماتی نشان دهند که هرگز در واقعیت رخ نداده است. قرار دادن افراد در رویدادهای سیاسی و مجرمانه از جمله صحنه‌های قتل و سرقت که در واقعیت رخ نداده و می‌تواند منجر به خسارت‌های جبران ناپذیری برای افراد گردد. در ابتدا شاید اینطور به نظر می‌رسید که این تصاویر و ویدیوهای جعلی با چشم انسان قابل تشخیص هستند. اما با صرف هزینه و پیشرفت در فرآیند تولید آنها، کیفیت جعلی بودن آنها نیز افزایش یافت. در این شرایط، متخصصین حوزه سعی کردند تا با تولید مدل‌های مناسب و دقیق بتوانند تصاویر تولید شده توسط دیپ‌فیک را تشخیص دهند [۱۵]. با در نظر گرفتن قابلیت‌های شبکه‌های عصبی کانولوشنی در پردازش تصویر و استخراج ویژگی، مدل‌هایی با استفاده از این شبکه‌ها جهت تمایز تصاویر تولید شده توسط دیپ‌فیک از تصاویر واقعی ارائه گردیده است [۱۶]. در این راستا، تعدادی از مدل‌های از قبل آموزش دیده شده که بر روی حجم زیادی از داده‌ها آموزش دیده‌اند، مورد استفاده قرار گرفته‌اند. با تنظیم دقیق این مدل‌ها و اضافه نمودن تعدادی لایه، بهبودهایی در راستای تشخیص دیپ‌فیک ایجاد گردیده است. به عنوان نمونه، یک مدل مبتنی بر یادگیری عمیق به منظور تشخیص دیپ‌فیک در فریم‌های ویدیویی در مطالعه [۱۷] ارائه گردیده است. در این راستا، رویکرد XGBoost مورد استفاده قرار گرفته است. نواحی صورت با استفاده از یک آشکارساز چهره مبتنی بر مدل YOLO [۱۸]، یک شبکه عصبی کانولوشنی و نیز روش‌های Inception ResNet از فریم‌های ویدیو استخراج می‌گردد. مجموعه داده‌های CelebDF و FaceForencics++ برای ساخت و آموزش مدل مورد استفاده قرار گرفته‌اند. مدل پیشنهادی در این مطالعه موفق به کسب دقت ۹۰ درصدی برای تشخیص جعلی بودن یا نبودن ویدیو دست یافته است. در مطالعه‌ای دیگر، مدل‌های MobileNet و Xception برای تشخیص جعلی بودن یا نبودن فریم‌های ویدیو مورد استفاده قرار گرفته است. نتایج به دست آمده بر روی پایگاه داده FaceForensics++ حاکی از دقت تشخیص در بازه‌ی ۹۱٪ تا ۹۸٪ می‌باشد [۱۹]. DeepVision، رویکرد پیشنهادی بر اساس شبکه‌های زایشی عمیق می‌باشد که برای تشخیص دیپ‌فیک بر اساس الگوهای پلک زدن چشم انسان در عکس‌های ورودی به کار گرفته می‌شود [۲۰]. نتایج مدل بر روی داده‌های جمع‌آوری شده حاکی از دقت ۸۷ درصدی برای تشخیص دیپ فیک می‌باشد. استفاده از ناسازگاری‌های طیفی، مکانی و زمانی با استفاده از روش‌های یادگیری عمیق با ورودی‌های مختلف، اساس رویکرد پیشنهادی در [۲۱] می‌باشد. در این راستا، یک شبکه چندوجهی بر اساس شبکه‌های حافظه بلند مدت کوتاه پیشنهاد شده است. مجموعه داده چالش عمیق فیس بوک برای آموزش و آزمایش مدل استفاده شده است که مدل پیشنهادی موفق به کسب دقت ۶۱ درصد برای تشخیص دیپ‌فیک گردیده است. استفاده از مدل‌های زایشی عمیق جهت تولید تصویر جعلی و واقعی در مدل‌های مبتنی بر ورودی‌های زوجی، یکی دیگر از رویکردهای پیشنهادی جهت تشخیص دیپ‌فیک می‌باشد [۲۲]. نتایج گزارش شده بر روی پایگاه داده CelebA حاکی از کسب دقت تشخیص ۹۰ درصدی برای دیپ‌فیک می‌باشد. رویکردهای یادگیری گروهی نیز جهت تشخیص فریم‌های جعلی ویدیو به کار گرفته شده‌اند [۲۳]. در این راستا، مدل DeepfakeStack ارائه شده است که متشکل از تعدادی مدل‌های یادگیری عمیق به صورت گروهی جهت تشخیص دیپ‌فیک در فریم های ویدیو می‌باشد. نتایج تجربی بر روی پایگاه داده

Archive of SID.ir

Archive of SID.ir

FaceForensics++ حاکی از بهبود نتایج می‌باشد.

بررسی مطالعات و مدل‌های ارائه شده نشان می‌دهد که این مدل‌ها در تشخیص برخی از تصاویر جعلی ناتوان هستند. علاوه بر این، مدل‌های ارائه شده قابلیت‌های متفاوتی در تشخیص انواع مختلفی از دیپ‌فیک‌ها دارند. در این مقاله، به منظور بهبود دقت تشخیص دیپ‌فیک در تصاویر، تعدادی از مدل‌های مبتنی بر شبکه‌های عصبی کانولوشنی با یکدیگر ترکیب گردیده‌اند تا امکان استفاده از قابلیت‌های این مدل‌ها به صورت همزمان ایجاد گردد.

در ادامه این مقاله، در بخش دوم، تعدادی از مدل‌های از قبل آموزش دیده شده به صورت مختصر معرفی می‌گردند. مدل پیشنهادی و نتایج به دست آمده در بخش‌های سه و چهار ارائه می‌گردند. در نهایت، بخش پنج به نتیجه‌گیری و پیشنهادات آینده اختصاص می‌یابد.

۲- شبکه‌های عصبی کانولوشنی از قبل آموزش دیده شده

در این بخش، تعدادی از مدل‌های شبکه‌های عصبی کانولوشنی از قبل آموزش دیده شده که در مدل پیشنهادی مورد استفاده قرار گرفته‌اند، به صورت مختصر معرفی می گردند.

DenseNet201 -۲-۱

DenseNet-201 یک شبکه عصبی کانولوشن است که ۲۰۱ لایه عمق دارد [۲۴]. این مدل بر روی بیش از یک میلیون تصویر از پایگاه داده ImageNet آموزش دیده است. این شبکه از پیش آموزش دیده می تواند تصاویر را به ۱۰۰۰ دسته طبقه‌بندی کند. شکل (۱)، معماری این مدل را نشان می‌دهد.

EfficientNetB2 -۲-۲

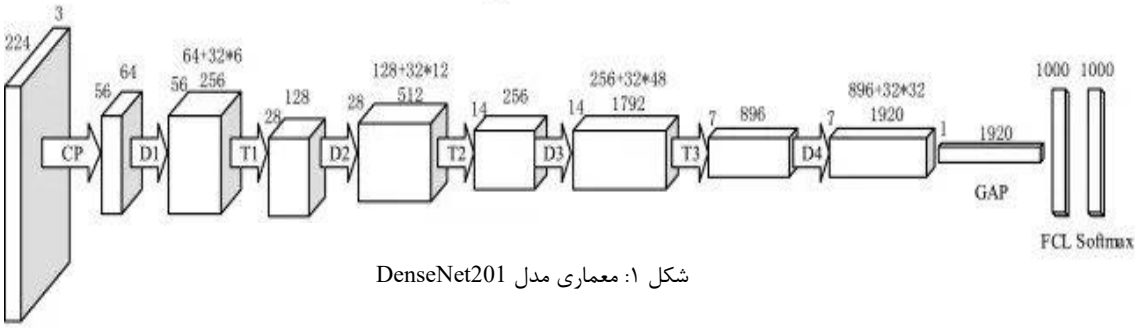
EfficientNetB2، که برای اولین بار در سال ۲۰۱۹ معرفی شد، یکی از کارآمدترین مدل‌ها برای دسته‌بندی تصاویر می‌باشد [۲۵]. این مدل، نسخه‌های مختلفی دارد که از B0 تا B7 می‌باشد. در این مقاله از نسخه B2 استفاده می‌شود که شامل ۳۴۲ لایه می‌باشد. شکل (۲) معماری این مدل را نشان می‌دهد.

Inception-ResNet-V2 -۲-۳

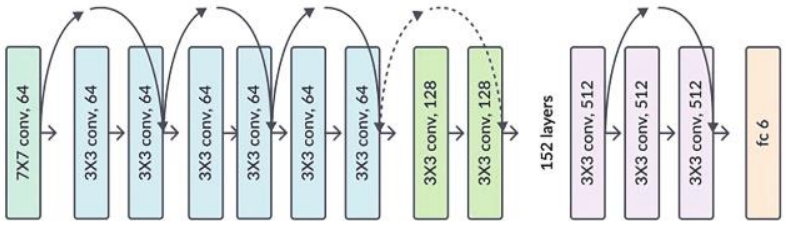
Inception-ResNet-V2 یک شبکه عصبی کانولوشنی است که بر اساس خانواده معماری‌های Inception ساخته شده است. اگرچه، اتصالات باقیمانده جایگزین مرحله الحاق فیلتر در معماری Inception شده است. این شبکه دارای ۱۶۴ لایه می‌باشد [۲۶]. شکل (۳) معماری این مدل را نشان می‌دهد.

ResNet152 -۲-۴

مدل ResNet152 در سال ۲۰۱۵ توسط Microsoft Research Asia منتشر شد، معماری ResNet152 نتایج بسیار موفقی را در ImageNet به دست آورده است [۲۷]. مدل ResNet152، دارای ۱۵۲ لایه می‌باشد. شکل (۴)، معماری این مدل را نشان می‌دهد.



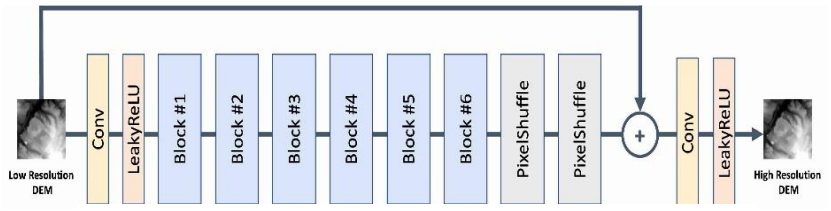
شکل ۱: معماری مدل DenseNet201



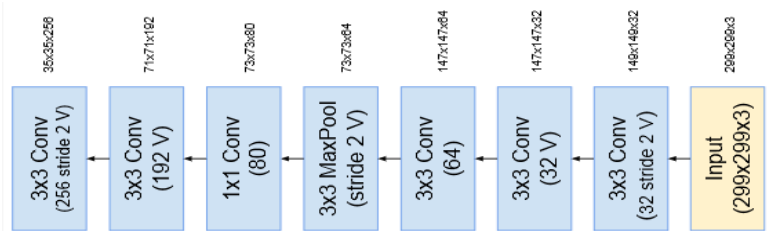
شکل ۲: معماری مدل ResNet152

Archive of SID.ir

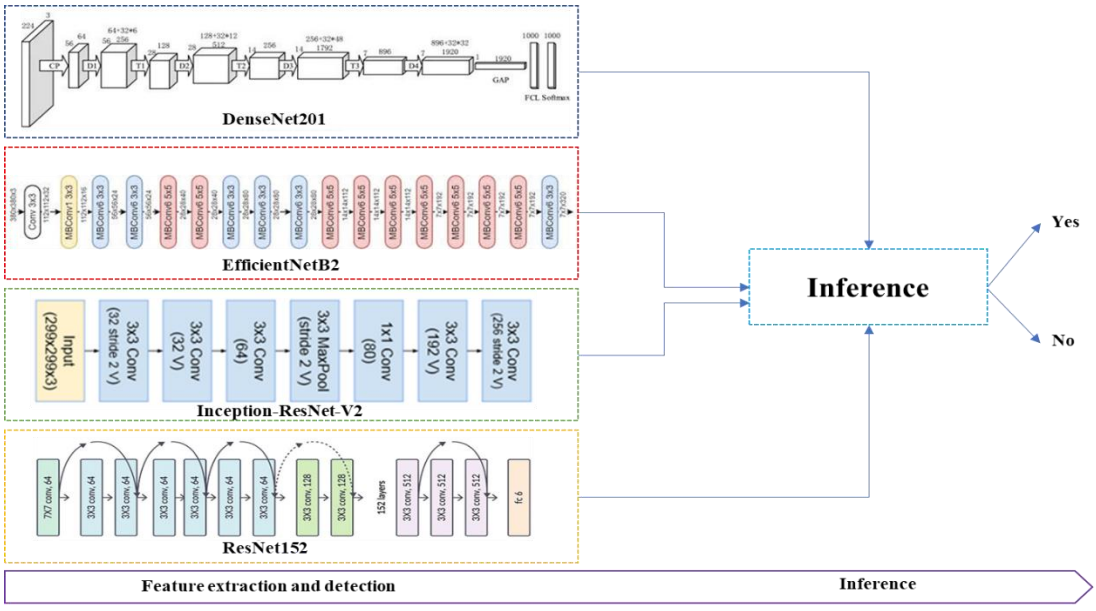
Archive of SID.ir



شکل ۴: معماری مدل EfficientNetB2



شکل ۳: معماری مدل Inception-ResNet-V2



شکل ۵: نمای کلی مدل پیشنهادی.

جدول ۱: مقایسه نتایج مدل ترکیبی پیشنهادی با بهترین مدل‌های موجود بر روی سه پایگاه داده.

مدل	140k Real And Fake Faces	DFDC Faces	Deepfake and Real Images
[۳۱]	۹۸.۰۰	-	-
[۳۲]	-	۹۵.۱۰	-
[۳۳]	-	-	۸۳.۰۶
مدل پیشنهادی	**۹۹.۱۲**	**۹۶.۲۴**	**۹۸.۸۰**

Archive of SID.ir

Archive of SID.ir

جدول ۲: نتایج به دست آمده از دقت تشخیص چهار شبکه عصبی کانولوشنی و مدل ترکیبی پیشنهادی بر روی سه پایگاه داده.

مدل	Deepfake and Real Images	DFDC Faces	140k Real and Fake Faces
DenseNet201	۹۶.۳۸	۸۳.۲۰	۸۹.۶۵
EfficientNetB2	۹۵.۵۱	۸۶.۳۰	۹۸.۵۷
ResNet152	۹۵.۵۸	۸۶.۵۲	۹۳.۶۰
InceptionResNetV2	۹۵.۹۱	۸۲.۷۸	۹۷.۷۹
مدل ترکیبی پیشنهادی	**۹۸.۸۰**	**۹۶.۲۴**	**۹۹.۱۲**
میزان بهبود نسبت به بهترین نتیجه موجود	**۲.۴۲**	**۹.۷۲**	**۰.۵۵**

جدول ۳: نتایج به دست آمده از تلفات تشخیص چهار شبکه عصبی کانولوشنی و مدل ترکیبی پیشنهادی بر روی سه پایگاه داده.

مدل	Deepfake And Real Images	DFDC Faces	140k Real And Fake Faces
DenceNet201	۰.۱۱	۰.۷۴	۰.۲۸
EfficientNetB2	۰.۱۲	۰.۴۹	۰.۰۴
ResNet152	۰.۱۲	۰.۴۱	۰.۱۹
InceptionResNetV2	۰.۱۰	۰.۸۹	۰.۰۶
مدل ترکیبی پیشنهادی	**۰.۰۴**	**۰.۲۸**	**۰.۰۲**
میزان بهبود نسبت به بهترین نتیجه موجود	**۰.۰۸**	**۰.۱۳**	**۰.۰۲**

۳- مدل پیشنهادی

در این بخش، جزئیات مدل پیشنهادی ارائه داده می‌شود. شکل ۵، نمای کلی مدل پیشنهادی را نشان می‌دهد.

۳-۱- استخراج ویژگی و تشخیص

این بخش از مدل شامل چهار شبکه عصبی کانولوشنی می باشد: DenseNet201، EfficientNetB2، Inception-ResNet-V2 و ResNet152. با تکیه بر قابلیت‌های بالای این شبکه‌ها در استخراج ویژگی‌های موثر از تصویر ورودی، این بخش از مدل قادر به تشخیص همزمان دیپ‌فیک بودن یا نبودن تصویر ورودی توسط این چهار مدل می‌باشد.

۳-۲- استنتاج گروهی

این بخش از مدل، مسئولیت استناج گروهی در رابطه با تشخیص نهایی مدل را بر عهده دارد. ورودی بخش استنتاج گروهی، یک بردار دودویی است که شامل چهار مقدار می‌باشد. این مقادیر، نتایج به دست آمده از تشخیص چهار مدل شبکه عصبی در بخش قبلی مدل می‌باشند. در صورتی که حداقل یکی از مقادیر این بردار برابر با ۱ باشد، نتیجه نهایی نیز برابر با ۱ خواهد بود. خروجی ۱ به معنای دیپ‌فیک بودن تصویر ورودی می‌باشد.

در صورتی که تمام مقادیر بردار دودویی، صفر باشند، نتیجه خروجی نیز صفر خواهد بود که به معنای واقعی بودن تصویر ورودی می‌باشد.

۵- نتایج تجربی

در این بخش، جزئیات پیاده‌سازی مدل، پایگاه داده‌های مورد استفاده، پیش‌پردازش‌های صورت گرفته بر روی داده ها و نیز نتایج ارزیابی مدل پیشنهادی ارائه می‌گردد.

Archive of SID.ir

Archive of SID.ir

۵-۱- جزئیات پیاده‌سازی مدل

پیاده‌سازی مدل پیشنهادی با استفاده از زبان برنامه‌نویسی پایتون و کتابخانه کراس صورت گرفته است. برای آموزش شبکه از پردازنده گرافیکی Tesla-K80 و سیستم عامل ویندوز ۱۰ استفاده گردیده است. در طی آموزش مدل، نرخ یادگیری برابر با ۰.۰۰۱ و اندازه دسته نیز ۶۴ در نظر گرفته شده است. با استفاده از روش Early Stopping، تعداد تکرار برابر با ۲۰۰ در نظر گرفته شده است.

۵-۲- پایگاه داده‌های مورد استفاده

برای آموزش و آزمایش مدل پیشنهادی، سه پایگاه داده مورد استفاده قرار گرفته است که جزئیات آنها در ادامه این بخش قابل مشاهده می‌باشد.

۵-۲-۱- دیتاست 140k Real and Fake Faces

پایگاه داده‌ی 140k Real and Fake Faces، حاوی تصاویر جعلی و واقعی استخراج شده از ویدیوهای پایگاه داده‌ی Deep Fake Detection Challenge می‌باشد [۲۸]. این پایگاه داده شامل ۷۰ هزار تصویر چهره واقعی و ۷۰ هزار تصویر چهره جعلی می‌باشد که هر تصویر دارای ابعاد ۲۵۶ در ۲۵۶ می‌باشد.

۵-۲-۲- پایگاه داده DFDC Faces

پایگاه داده‌ی DFDC Faces شامل ۲۰۷۰۰ تصویر چهره واقعی و ۷۳۲۰۰ تصویر چهره جعلی می‌باشد. این پایگاه داده، تصاویر جعلی را از ویدیوهای جعلی استخراج کرده است و فریم‌های آن را به صورت منسجم ارائه کرده است. ابعاد هر تصویر، ۲۵۶ در ۲۵۶ می‌باشد [۲۹].

۵-۲-۳- پایگاه داده Deepfake and Real Images

پایگاه داده‌ی Deepfake and Real Images شامل ۷۰۰۰۰ تصویر چهره واقعی و ۷۰۰۰۰ تصویر چهره جعلی برای داده های آموزشی، تعداد ۱۹۶۰۰ تصویر چهره واقعی و ۱۹۸۰۰ تصویر چهره جعلی برای داده‌های ارزیابی، ۵۴۱۳ تصویر چهره واقعی و ۵۴۹۲ تصویر چهره جعلی برای داده-های آزمون می‌باشد. ابعاد هر تصویر، ۲۵۶ در ۲۵۶ می‌باشد [۳۰].

۵-۳- پیش‌پردازش داده‌ها

به منظور استفاده از مدل‌های از قبل آموزش دیده شده، لازم است تا ابعاد تصویر ورودی با ابعاد لایه ورودی این مدل‌ها سازگار باشد. برای این منظور، تصاویر پایگاه داده-های مورد استفاده نیز تغییر اندازه داده می‌شوند تا قابل استفاده در این مدل‌ها باشند.

۵-۴- نتایج ارزیابی

در این بخش، نتایج ارزیابی مدل پیشنهادی بر روی سه پایگاه داده نشان ارائه می‌گردد. برای این منظور، در ابتدا، نتایج به دست آمده از بخش‌های مختلف مدل نشان داده می‌شود. پس از آن، نتایج مربوط به کل مدل ارائه می‌گردد.

۵-۴-۱- نتایج مدل DenseNet201

دقت این مدل بر روی پایگاه داده 140k Real And Fake Faces برابر با ۸۹.۶۵ درصد، بر روی پایگاه داده DFDC Faces برابر با ۸۳.۲۰ درصد و بر روی پایگاه داده Deepfake and Real Images برابر با ۹۶.۳۸ درصد می باشد. نمودارهای مربوط به دقت و فقدان شبکه DenseNet201 بر روی سه پایگاه داده نشان می‌دهد که شبکه DenseNet201 پس از طی تعدادی تکرار، به وضعیت پایداری می‌رسد.

۵-۴-۲- نتایج مدل EfficientNetB2

دقت مدل EfficientNetB2 بر روی پایگاه داده 140k Real And Fake Faces برابر با ۹۸.۵۷ درصد، برروی پایگاه داده DFDC Faces برابر با ۸۶.۳۰ درصد و بر روی پایگاه داده Deepfake and Real Images برابر با ۹۵.۵۱ درصد می باشد. نمودارهای مربوط به دقت و فقدان شبکه EfficientNetB2 بر روی سه پایگاه داده نشان می‌دهد که شبکه EfficientNetB2 پس از طی تعدادی تکرار، به وضعیت پایداری می‌رسد.

۵-۴-۳-نتایج مدل ResNet152

دقت مدل ResNet152 بر روی پایگاه داده 140k Real and Fake Faces برابر با ۹۳.۶۰ درصد، بر روی پایگاه داده DFDC Faces برابر با ۸۶.۵۲ درصد و بر روی پایگاه داده Deepfake and Real Images برابر با ۹۵.۵۸ درصد می باشد. نمودارهای مربوط به دقت و فقدان شبکه ResNet152 بر روی سه پایگاه داده نشان می‌دهد که شبکه ResNet152 پس از طی تعدادی تکرار، به وضعیت پایداری می‌رسد.

۵-۴-۴- نتایج مدل InceptionResNetV2

دقت مدل InceptionResNetV2 بر روی پایگاه داده 140k Real And Fake Faces برابر با ۹۷.۷۹ درصد، بر روی پایگاه داده DFDC Faces برابر با ۸۲.۹۸ درصد و بر روی پایگاه داده Deepfake and Real Images برابر با ۹۵.۹۱ درصد می‌باشد. نمودارهای مربوط به دقت و فقدان شبکه ResNet152 بر روی سه پایگاه داده نشان می‌دهد که شبکه

Archive of SID.ir

Archive of SID.ir

ResNet152 پس از طی تعدادی تکرار، به وضعیت پایداری می‌رسد.

۵-۴-۵- نتایج مدل ترکیبی

دقت مدل ترکیبی بر روی پایگاه داده 140k Real And Fake Faces برابر با ۹۹.۱۲ درصد، بر روی پایگاه داده DFDC faces برابر با ۹۶.۲۴ درصد و بر روی پایگاه داده Deepfake and Real Images برابر با ۹۸.۸۰ درصد می‌باشد. نمودارهای مربوط به دقت و فقدان مدل پیشنهادی بر روی سه پایگاه داده نشان می‌دهد که مدل پیشنهادی پس از طی تعدادی تکرار، به وضعیت پایداری می‌رسد.

۵-۴-۶- بحث بر روی نتایج

بررسی نتایج به دست آمده از چهار شبکه عصبی کانولوشنی DenseNet201، EfficientNetB2، Inception-ResNet-V2 و ResNet152 و مقایسه آنها با مدل ترکیبی نشان می‌دهد که با تکیه بر قابلیت‌های بالای این شبکه‌ها در استخراج ویژگی‌های موثر از تصویر ورودی، مدل ترکیبی قادر به تشخیص بهتر و همزمان دیپ‌فیک بودن یا نبودن تصویر ورودی توسط این چهار مدل می‌باشد. جدول ۱، مقایسه نتایج به دست آمده از مدل پیشنهادی با بهترین نتایج موجود بر روی سه پایگاه داده را نشان می‌دهد. مقایسه مدل ترکیبی پیشنهادی با بهترین نتایج موجود بر روی سه پایگاه داده مورد استفاده حاکی از بهبود نتایج می‌باشد.

علاوه بر این، جدول ۲ و ۳ نتایج مربوط به دقت و فقدان شبکه‌های عصبی کانولوشنی و نیز مدل ترکیبی را بر روی سه پایگاه داده نشان می‌دهد. همانگونه که این جدول‌ها نشان می‌دهند، مدل ترکیبی سبب بهبود عملکرد تشخیص دیپ‌فیک در تصویر ورودی گردیده است.

۶- نتیجه‌گیری و کارهای آینده

در این مقاله، یک مدل ترکیبی با استفاده از چهار شبکه عصبی کانولوشنیِ DenseNet201، EfficientNetB2، Inception-ResNet-V2 و ResNet152 ارائه گردید. با تکیه بر قابلیت‌های بالای این شبکه‌ها در استخراج ویژگی‌های موثر از تصویر ورودی، مدل پیشنهادی قادر به تشخیص همزمان دیپ‌فیک بودن یا نبودن تصویر ورودی توسط این چهار مدل می‌باشد. نتایج ارائه شده بر روی سه پایگاه داده 140k Real and Fake Faces، DFDC و Deepfake and Real Images حاکی از بهبود نتایج نسبت به مدل‌های موجود می‌باشد. علاوه بر این، نتایج مربوط به دقت و فقدان شبکه‌های عصبی کانولوشنی و نیز مدل ترکیبی ارائه گردید و مورد بحث قرار گرفته است. در راستای بهبود دقت مدل، در کارهای آینده، قابلیت‌های ترنسفورمر بینایی مورد توجه قرار خواهد گرفت. علاوه بر این، توالی تصاویر ورودی نیز مورد بررسی قرار خواهد گرفت.

مراجع

[1] Zobaed, Sm, and Md Fazle Rabby, Md Istiaq Hossain, Ekram Hossain, Sazib Hasan, Asif Karim, and Khan Md. Hasib. "Deepfakes: Detecting forged and synthetic media content using machine learning". *Artificial Intelligence in Cyber Security: Impact and Implications* (2021): 177–201.

[2] Rastgoo, Razieh and Vahid Sattari Naeini. "A neurofuzzy QoS-aware routing protocol for smart grids". 22nd Iranian Conference on Electrical Engineering (ICEE), pp. 1080-1084, 2014.

[3] Rastgoo, Razieh and Vahid Sattari Naeini. "Tuning parameters of the QoS-aware routing protocol for smart grids using genetic algorithm". *Applied Artificial Intelligence* 30, no. 1 (2016): 52-76.

[4] Majidi, Neza, Kourosh Kiani, and Razieh Rastgoo. "A deep model for super-resolution enhancement from a single image". *Journal of AI and Data Mining* 8, no. 4, (2020): 451-460.

[5] Kiani, Kourosh, Razieh Hematpour, and Razieh Rastgoo. "Automatic grayscale image colorization using a deep hybrid model". *Journal of AI and Data Mining* 9, no. 3 (2021): 321-328.

[6] Rastgoo, Razieh and Vahid Sattari-Naeini. "Gsomcr: Multi-constraint genetic-optimized qos-aware routing protocol for smart grids". *Iranian Journal of Science and Technology, Transactions of Electrical Engineering* 42, (2018): 185-194.

[7] Rastgoo, Razieh and Kourosh Kiani. "Face recognition using fine-tuning of Deep Convolutional Neural Network and transfer learning". *Journal of Modeling in Engineering* 17, no. 58 (2019): 103-111.

[8] Rastgoo, Razieh, Kourosh Kiani, Sergio Escalera, and Mohammad Sabokrou. "Multi-modal zero-shot sign language recognition". *arXiv:2109.00796*, (2021).

Archive of SID.ir

Archive of SID.ir

[9] Zarbafi, Sahar, Kourosh Kiani, and Razieh Rastgoo. "Spoken Persian digits recognition using deep learning". *Journal of Modeling in Engineering* 21, (2023): 163-172.

[10] Alinezhad, Fatemeh, Kourosh Kiani, and Razieh Rastgoo. "A Deep Learning-based Model for Gender Recognition in Mobile Devices". *Journal of AI and Data Mining* 11, (2023): 229-236.

[11] Rastgoo, Razieh, Kourosh Kiani, and Sergio Escalera. "ZS-SLR: Zero-Shot Sign Language Recognition from RGB-D Videos". *arXiv:2108.10059*, (2021).

[12] Thambawita, Vajira, and et al. "DeepFake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine". *Sci. Rep.* 11, (2021): 21869.

[13] Faisal Bin Ahmed, Mohammad, M. Saef Ullah Miah, Abhijit Bhowmik, and Juniada Binti Sulaiman. "Awareness to Deepfake: A resistance mechanism to Deepfake". In Proceedings of the 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen, pp. 1–5, 2021.

[14] Gautam, Neil, and Dinesh Kumar Vishwakarma. "Obscenity Detection in Videos through a Sequential ConvNet Pipeline Classifie". *IEEE Trans. Cogn. Dev. Syst.* 15, no. 1 (2023): 310-318.

[15] Naik, Rakesh. "Deepfake Crimes: How Real and Dangerous They Are in 2021? ". Available online: https://cooltechzone.com/research/deepfake-crimes. Accessed Date: Jan 2024.

[16] Jin, Bo, Leandro Cruz, and Nuno Gonçalves. "Deep facial diagnosis: Deep transfer learning from face recognition to facial diagnosis". *IEEE Access* 8, (2020): 123649–123661.

[17] Ismail, Aya, Marwa Elpeltagy, Mervat S. Zaki, and Kamal Eldahshan. "A New Deep Learning-Based Methodology for Video Deepfake Detection Using XGBoost". *Sensors* 21, (2021): 5413.

[18] Chen, Weijun, Hongbo Huang, Shuai Peng, Changsheng Zhou, and Cuiping Zhang. "YOLO-face: A real-time face detector". *Vis. Comput.* 37, (2021): 805–813.

[19] Pan, Deng, Lixian Sun, Rui Wang, Xingjian Zhang, and Richard O. Sinnott. "Deepfake Detection through Deep Learning". In Proceedings of the 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), Leicester, UK, pp. 134–143, 2020.

[20] Jung, Tackhyun, Sangwon Kim, and Keecheon Kim. "DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern". *IEEE Access* 8, (2020): 83144-83154.

[21] Lewis, John K., Imad Eddine Toubal, Helen Chen, Vishal Sandesera, Michael Lomnitz, Zigfried Hampel-Arias, Calyam Prasad, and Kannappan Palaniappan. "Deepfake Video Detection Based on Spatial, Spectral, and Temporal Inconsistencies Using Multimodal Deep Learning". In Proceedings of the 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington DC, USA, pp. 1–9, 2020.

[22] Hsu, Chih-Chung, Yi-Xiu Zhuang, and Chia-Yen Lee. "Deep fake image detection based on pairwise learning". *Appl. Sci.* 10, (2020): 370.

[23] Rana, Md. Shohel, and Andrew H. Sung. "DeepfakeStack: A Deep Ensemble-based Learning Technique for Deepfake Detection". In Proceedings of the 2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom), New York, NY, USA, pp. 70–75, 2020.

[24] Huang, Gao, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. "Densely Connected Convolutional Networks". In Proceedings of the 2017 Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, 2017.

[25] Tan, Mingxing, and Quoc V. Le. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks". In Proceedings of the 2019 International Conference on Machine Learning (ICML), Jun 2019.

[26] Szegedy, Christian, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning". In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17), pp. 4278–4284, 2017.

[27] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep Residual Learning for Image Recognition". In Proceedings of the 2017 Computer Vision and Pattern Recognition (CVPR), Caesars Palace, 2016.

[28] https://www.kaggle.com/datasets/xhlulu/140k-real-and-fake-faces. Access Date: Jan. 2024.

[29] https://www.kaggle.com/datasets/itamargr/dfdc-faces-of-the-train-sample, Access Date: Jan. 2024.

[30] https://www.kaggle.com/datasets/manjilkarki/deepfake-and-real-imagesv, Access Date: Jan. 2024.

Archive of SID.ir

Archive of SID.ir

[31] Raza, Ali, Kashif Munir, and Mubarak Almutairi, "A Novel Deep Learning Approach for Deepfake Image Detection". *Appl. Sci.* 12,(2022): 9820.

[32] https://www.kaggle.com/code/abdalrhmanmorsi/real-vs-fake-face-cnn-model, Access Date: Jan. 2024.

[33] https://www.kaggle.com/code/shubhanshu609/deepfake-cnn-2, Access Date: Jan. 2024.

Archive of SID.ir