Analýza mikroexpresií pre detekciu deep fake videí

Microexpression analysis for deep fake video detection

Peter Procházka[1]

Abstrakt
Technológia deep fake, využívajúca pokročilé neurónové siete, predstavuje významnú výzvu v oblasti informačnej bezpečnosti a autenticity digitálnych médií. Tento článok skúma možnosť využitia mikroexpresií, krátkych a nevedomých zmien vo výraze tváre, na detekciu deep fake videí. Naša metodológia zahŕňa detekciu mikroexpresií v reálnych videách pomocou konvolučných neurónových sietí a následnú analýzu rozdielov medzi reálnymi a syntetickými videami. Výsledky ukazujú, že deep fake videá majú výrazné ťažkosti s reprodukciou mikroexpresií, čo vedie k nižšej frekvencii a kratšiemu trvaniu týchto výrazov. Na základe týchto zistení vyvíjame algoritmus, ktorý by efektívne identifikoval deep fake videá s čo možno najvyššou presnosťou. Tento výskum ponúka nový prístup k detekcii syntetických videí a podčiarkuje význam mikroexpresií pri zabezpečení autenticity digitálnych médií.

Kľúčové slová
mikroexpresie, deep fake videá, detekcia, neurónové siete

Abstract
Deep fake technology, utilizing advanced neural networks, presents a significant challenge in the realm of information security and digital media authenticity. This paper explores the potential of using microexpressions, brief and unconscious changes in facial expressions, to detect deep fake videos. Our methodology involves detecting microexpressions in real videos using convolutional neural networks and subsequently analyzing the differences between real and synthetic videos. The results indicate that deep fake videos have significant difficulties reproducing microexpressions, leading to lower frequency and shorter duration of these expressions. Based on these findings, we are developing an algorithm that would effectively identify deep fake videos with the highest possible accuracy. This research offers a novel approach to detecting synthetic videos and underscores the importance of microexpressions in ensuring the authenticity of digital media.

Key words
microexpressions, deep fake videos, detection, neural networks

JEL classification
C63, O33

1 Úvod

V posledných rokoch sa technológia deep fake stala jednou z najdiskutovanejších tém v oblasti informačnej bezpečnosti a digitálnej etiky. Deep fake videá, ktoré sú vytvárané pomocou pokročilých algoritmov hlbokého učenia, sú schopné generovať realistické falošné videá, ktoré môžu byť použité na rôzne účely - od zábavy až po vážnejšie hrozby, ako sú politická manipulácia, podvody a podobne.

[1] Ekonomická univerzita v Bratislave, Fakulta hospodárskej informatiky, Katedra aplikovanej informatiky, Dolnozemská cesta 1, Bratislava, Slovakia, peter.prochazka@euba.sk

Jedným z hlavných problémov spojených s deep fake videami je ich potenciál na šírenie dezinformácií. Vzhľadom na to, že tieto videá môžu byť veľmi realistické, môže byť pre divákov ťažké rozoznať, čo je pravdivé a čo je falošné. To môže viesť k šíreniu falošných správ a manipulácii verejnej mienky. Navyše, deep fake videá môžu byť použité na vydieranie, poškodzovanie reputácie alebo na vytváranie falošných dôkazov v právnych prípadoch.

Ďalším aspektom, ktorý zvyšuje naliehavosť riešenia problému deep fake videí, je ich dostupnosť. V súčasnosti existuje množstvo nástrojov a softvérov, ktoré umožňujú aj laikom vytvárať deep fake videá s relatívne vysokou kvalitou. To znamená, že riziko zneužitia tejto technológie je vysoké a môže mať ďalekosiahle následky pre jednotlivcov aj spoločnosť.

Práve tieto potenciálne zneužitia deep fake technológie vytvárajú naliehavú potrebu vyvinúť účinné metódy na detekciu a overovanie autenticity multimediálnych materiálov.

Významným faktorom v hodnotení autentickosti videí môžu byť mikroexpresie. Mikroexpresie, ako krátke a nevedomé prejavy emócií, poskytujú jedinečný pohľad na skutočný emocionálny stav človeka. Tieto jemné zmeny vo výraze tváre sú výsledkom rýchlych svalových pohybov a trvajú len niekoľko milisekúnd. Vďaka svojej krátkej trvanlivosti a jemným detailom sú mikroexpresie ťažko reprodukovateľné aj pre najpokročilejšie deep fake technológie. Analýza týchto mikroexpresií preto predstavuje sľubnú metódu na odhalenie syntetických videí, ktoré môžu inak pôsobiť veľmi realisticky.

Cieľom tohto článku je podrobne preskúmať možnosti využitia mikroexpresií na detekciu deep fake videí. V prvej časti článku sa zameriame na teoretické pozadie deep fake technológií a mikroexpresií. Následne budeme diskutovať o cieľoch nášho výskumu a metodológii, ktorú sme použili na detekciu mikroexpresií a analýzu ich rozdielov medzi reálnymi a syntetickými videami. V experimentálnej časti článku predstavíme výsledky našich testov a vyhodnotíme výkonnosť vyvinutého detekčného algoritmu. Nakoniec budeme diskutovať o interpretácii výsledkov, porovnaní s existujúcimi metódami, obmedzeniach našej štúdie a potenciálnych budúcich smerovaniach výskumu.

2 Súčasný stav a teoretické pozadie

Existuje niekoľko nástrojov a softvérových platforiem, ktoré umožňujú aj laikom jednoducho generovať deep fake videá. Tieto nástroje sú často dostupné online a ponúkajú rôzne úrovne funkcií a prispôsobenia. Medzi najznámejšie patria:

1. FaceApp: Táto aplikácia je široko používaná na úpravu fotografií a videí. Používa neurónové siete na transformáciu tvárí, čo zahŕňa zmeny veku, pohlavia, účesu a ďalších atribútov. Hoci je primárne určená na zábavné účely, technológia použitá v aplikácii je veľmi pokročilá.
2. Reface: Je ďalšia populárna aplikácia, ktorá umožňuje používateľom nahradiť tváre v krátkych videách a GIF-och. Aplikácia využíva technológiu syntézy obrazu na vytváranie veľmi realistických deep fake videí v reálnom čase.
3. DeepFaceLab: Tento nástroj je jedným z najpokročilejších a najrozšírenejších v komunite tvorcov deep fake videí. Umožňuje detailné úpravy a je široko využívaný na akademické aj neakademické účely. Používateľom umožňuje vytvárať a trénovať vlastné modely na syntézu tvárí.
4. Faceswap: Je open-source projekt, ktorý je podobný DeepFaceLab. Tento nástroj je prístupný pre používateľov všetkých úrovní znalostí a poskytuje detailné návody a podporu pre vytváranie deep fake videí. Umožňuje výmenu tvárí vo videách s vysokou presnosťou.
5. Zao: Je čínska aplikácia, ktorá sa stala veľmi populárnou vďaka svojej schopnosti generovať realistické deep fake videá v priebehu niekoľkých sekúnd. Používatelia môžu nahrávať svoje fotografie a aplikácia automaticky vytvorí video s ich tvárou na mieste známych hercov alebo postáv.

6. MyHeritage Deep Nostalgia: Táto služba je špeciálne navrhnutá na oživenie starých rodinných fotografií. Používa technológiu na animáciu tvárí na fotografiách, čo vytvára dojem, že osoby na fotografiách sa pohybujú a usmievajú. Hoci je primárne určená na genealogické účely, technológia za ňou je veľmi pokročilá a môže byť zneužitá na tvorbu deep fake videí.
7. Avatarify: Umožňuje používateľom používať živé deep fake technológie na videokonferencie. Tento nástroj môže byť použitý na transformáciu tvárí v reálnom čase, čo umožňuje používateľom napríklad predstierať, že sú niekým iným počas videohovorov.
8. Deep Art Effects: Tento nástroj umožňuje používateľom aplikovať umelecké štýly na fotografie a videá, pričom využíva technológie hlbokého učenia. Hoci je primárne zameraný na kreatívne aplikácie, môže byť použitý aj na generovanie deep fake videí s vysokou mierou realizmu.
9. AvengeThem: Táto webová aplikácia umožňuje používateľom nahrávať svoje fotografie a vytvárať videá, kde sa ich tvár objavuje na postavách z populárnych filmov, najmä z Marvel Cinematic Universe. Hoci je aplikácia vytvorená na zábavné účely, používa technológie na výmenu tvárí, ktoré môžu byť zneužité na vytváranie deep fake videí.
10. Synthesia: Je nástroj, ktorý umožňuje tvorbu videí, kde avatar alebo osoba na videu hovorí text, ktorý používateľ zadá. Táto technológia je veľmi užitočná na tvorbu marketingových a edukačných videí, ale môže byť zneužitá na vytváranie falošných výpovedí alebo dezinformačných videí.
11. JigSaw: Je experimentálny nástroj od spoločnosti Google, ktorý umožňuje vytvárať deep fake videá na účely výskumu a vývoja detekčných metód. Hoci je jeho dostupnosť obmedzená, poskytuje výkonné nástroje na generovanie syntetických médií.
12. FakeApp: Bola jednou z prvých aplikácií na tvorbu deep fake videí, ktorá sa stala populárnou. Umožňuje používateľom vytvárať deep fake videá pomocou priateľského používateľského rozhrania. Aplikácia vyžaduje základné technické znalosti, ale poskytuje vysokú kvalitu výstupov.
13. Deep Video Portraits: Tento nástroj umožňuje upravovať výrazy tváre a pohyby hlavy v existujúcich videách. Používa technológie hlbokého učenia na vytváranie realistických animácií tváre, čo môže byť využité na tvorbu deep fake videí.
14. Deepfakes web β: Je to online platforma, ktorá umožňuje používateľom vytvárať deep fake videá bez potreby sťahovania softvéru. Táto platforma je navrhnutá tak, aby bola prístupná pre širokú verejnosť, a poskytuje relatívne jednoduché rozhranie na tvorbu falošných videí.
15. DeepFaceKit: Tento nástroj je navrhnutý pre pokročilých používateľov a výskumníkov, ktorí chcú vytvárať a študovať deep fake videá. Je to open-source projekt, ktorý ponúka množstvo funkcií na detailné úpravy a tréning modelov.
16. SimSwap: Je ďalší pokročilý nástroj, ktorý umožňuje používateľom vymieňať tváre vo videách s vysokou presnosťou. Tento nástroj je využívaný najmä v akademickom výskume a pri vývoji nových technológií na detekciu deep fake videí.

Použitie týchto nástrojov a softvérových platforiem predstavuje významné riziko pre šírenie dezinformácií a potrebu vyvinúť efektívne detekčné metódy na overovanie autenticity digitálnych médií. Je nevyhnutné, aby sa odborníci na informačnú bezpečnosť, vývojári a vedci zamerali na vytváranie a zdokonaľovanie technológií, ktoré budú schopné identifikovať a zabrániť šíreniu deep fake videí.

Deep fake technológie využívajú generatívne protivnícke siete (Generative Adversarial Networks, GANs) na vytváranie realistických falošných videí (Nguyen et al., 2019). Tieto siete pozostávajú z dvoch komponentov: generátora a diskriminátora (Goodfellow et al., 2014). Generátor vytvára syntetické obrazy, zatiaľ čo diskriminátor sa snaží odlíšiť tieto syntetické obrazy od skutočných. Tieto dve siete sa navzájom trénujú v protivníckom nastavení, čím generátor postupne zlepšuje kvalitu svojich výstupov.

Generatívne protivnícke siete sa ukázali byť mimoriadne efektívne pri vytváraní realistických obrazov a videí, pretože generátor sa neustále učí vytvárať obrazy, ktoré sú čoraz ťažšie odlíšiteľné od skutočných. Táto technológia bola pôvodne vyvinutá pre legitímne aplikácie, ako je tvorba realistických herných postáv a zlepšovanie kvality obrazov, avšak rýchlo sa rozšírila aj do menej etických oblastí.

S pokrokom v deep fake technológiách sa stali tieto videá čoraz sofistikovanejšími a ťažšie detekovateľnými. Napriek tomu však existujú určité obmedzenia. Napríklad, deep fake algoritmy často zápasia s reprodukciou jemných detailov, ako sú prirodzené pohyby očí, rýchle zmeny vo výraze tváre (mikroexpresie) a synchronizácia zvuku a obrazu. Tieto obmedzenia môžu byť využité na detekciu falošných videí.

Mikroexpresie: definícia a význam

Ako sme už uviedli, mikroexpresie sú krátke, nevedomé zmeny vo výraze tváre, ktoré trvajú len niekoľko milisekúnd. Sú výsledkom rýchlych a nevedomých pohybov svalov tváre, ktoré odrážajú skutočné emocionálne stavy človeka. Mikroexpresie boli prvýkrát detailne opísané psychológom Paulom Ekmanom a jeho kolegami, ktorí vyvinuli systém Facial Action Coding System (FACS) na analýzu a kódovanie týchto výrazov (Ekman & Friesen, 1978).

Mikroexpresie sú dôležité, pretože sú považované za autentické prejavy skutočných emócií. Zatiaľ čo makroexpresie, ktoré trvajú dlhšie, môžu byť ľahko kontrolované a manipulované, mikroexpresie sú príliš rýchle na vedomé ovládanie a preto poskytujú presnejší obraz o emocionálnom stave človeka. Tieto jemné zmeny vo výraze tváre môžu byť kľúčové pre detekciu nepravdivých výpovedí a identifikáciu skutočných emocionálnych reakcií.

V kontexte deep fake videí sú mikroexpresie zaujímavé tým, že ich reprodukcia je pre algoritmy hlbokého učenia veľmi náročná. Reálne mikroexpresie sú výsledkom zložitých fyziologických procesov a jemných svalových pohybov, ktoré sú ťažko simulovateľné. Preto môže analýza mikroexpresií poskytnúť účinný nástroj na odhalenie deep fake videí. Aj podľa (Ekman, 2009) deep fake videá majú výrazné ťažkosti s reprodukciou mikroexpresií.

Prehľad súčasných metód detekcie deep fake videí

Súčasné metódy detekcie deep fake videí sa zameriavajú na rôzne aspekty syntetických médií. Korshunov a Marcel (2018) diskutujú o rastúcej hrozbe technológie deep fake pre systémy rozpoznávania tváre a navrhujú rôzne detekčné techniky. Nguyen et al. (2019) poskytujú komplexný prehľad o tvorbe a detekcii deep fake pomocou metodológií hlbokého učenia. Tariq et al. (2018) sa zameriavajú na detekciu falošných obrazov tvárí vytvorených strojom aj človekom v reálnom prostredí, pričom predstavujú inovatívne prístupy na zvýšenie presnosti detekcie. Uvedené tímy autorov sa venujú napríklad analýze vizuálnych a zvukových artefaktov (Korshunov & Marcel, 2018), detekcii nesynchronizácie medzi obrazom a zvukom (Tariq et al., 2018), a využitiu strojového učenia na identifikáciu syntetických vzorov (Nguyen et al., 2019).

Niektoré z najbežnejších techník zahŕňajú:

1. Analýzu vizuálnych artefaktov:
   - Deep fake videá často obsahujú vizuálne chyby, ako sú nepresnosti v rozlíšení, deformácie tváre a neadekvátne osvetlenie. Tieto artefakty môžu byť identifikované algoritmami na analýzu obrazu.
   - Techniky na detekciu neprirodzených pohybov očí a tváre, môžu odhaliť nekonzistentné pohyby.
2. Detekciu anomálií vo zvuku:
   - Zvukové stopy v deep fake videách môžu vykazovať neautentické frekvenčné a amplitúdové charakteristiky. Analýza zvuku teda môže pomôcť identifikovať tieto nezrovnalosti.
   - Algoritmy tiež môžu analyzovať časovú synchronizáciu medzi obrazom a zvukom a zistiť nesúlad.
3. Využitie strojového učenia:
   - Hlboké neurónové siete môžu byť trénované na veľkých datasetoch obsahujúcich reálne a deep fake videá. Tieto modely sa učia rozlišovať medzi reálnymi a syntetickými videami na základe jemných rozdielov.
   - Transfer learning, kde sa modely trénované na súvisiacich úlohách adaptujú na detekciu deep fake videí, môže zvýšiť efektivitu a presnosť detekcie.
4. Kombinované prístupy:
   - Niektoré pokročilé prístupy kombinujú viacero techník na zvýšenie presnosti detekcie. Napríklad, kombinácia vizuálnej a zvukovej analýzy môže poskytnúť robustnejšie výsledky.

Každá z týchto metód má svoje výhody a nevýhody. Vizuálna analýza môže byť veľmi efektívna, ale je citlivá na vizuálne efekty a úpravy. Tieto môžu byť zámerne pridané do videa, aby zamaskovali stopy manipulácie. Napríklad, rôzne filtre, efekty a korekcie farieb môžu skryť nedokonalosti alebo artefakty, ktoré by inak mohli odhaliť, že video bolo falošné. Analýza zvuku je užitočná, ale môže byť obmedzená kvalitou zvukovej stopy. Metódy strojového učenia sú veľmi silné, ale vyžadujú veľké množstvo dát na tréning. Kombinované prístupy môžu ponúknuť najlepšie výsledky, ale sú tiež najkomplexnejšie a najnáročnejšie na implementáciu.

3 Ciele výskumu

Cieľom prvého kroku výskumu je vyvinúť spoľahlivé metódy na detekciu mikroexpresií v reálnych videách. Tieto metódy musia byť schopné zachytiť rýchle a jemné zmeny vo výraze tváre, ktoré trvajú len niekoľko milisekúnd. K tomu budeme potrebovať pokročilé techniky analýzy obrazu a strojového učenia, ktoré dokážu identifikovať a klasifikovať mikroexpresie na základe ich charakteristických znakov.

Pre tento účel plánujeme využiť konvolučné neurónové siete (CNN), ktoré sa ukázali byť mimoriadne efektívne pri analýze obrazových dát. Tieto siete budú trénované na rozsiahlych datasetoch obsahujúcich reálne videá s označenými mikroexpresiami. Cieľom je vytvoriť model, ktorý dokáže presne detekovať mikroexpresie v rôznych podmienkach a situáciách.

Po úspešnej identifikácii mikroexpresií v reálnych videách budeme skúmať rozdiely medzi mikroexpresiami v reálnych a deep fake videách. Tento krok zahŕňa štatistickú analýzu frekvencie, trvania a typu mikroexpresií v oboch kategóriách videí. Našim cieľom je identifikovať charakteristické znaky, ktoré sú prítomné v reálnych videách, ale chýbajú alebo sú nesprávne reprodukované v deep fake videách.

Na túto analýzu budeme používať rôzne štatistické metódy a techniky vizualizácie dát, aby sme mohli presne kvantifikovať a interpretovať zistené rozdiely. Predpokladáme, že deep

fake videá budú vykazovať nižšiu frekvenciu a kratšie trvanie mikroexpresií, ako aj vyššiu mieru nepresností a artefaktov v týchto výrazoch.

Na základe zistených rozdielov vyvinieme algoritmus, ktorý bude schopný efektívne detekovať deep fake videá pomocou analýzy mikroexpresií. Tento algoritmus bude kombinovať techniky strojového učenia a štatistickej analýzy na identifikáciu syntetických videí.

Algoritmus bude trénovaný na datasetoch obsahujúcich reálne a deep fake videá s označenými mikroexpresiami. Jeho výkon bude testovaný na nezávislej testovacej sade, aby sa overila jeho presnosť, precíznosť, a schopnosť detekovať deep fake videá v rôznych podmienkach. Cieľom je vytvoriť robustný a spoľahlivý nástroj na detekciu falošných videí, ktorý bude schopný pracovať v reálnom čase a poskytovať vysoko presné výsledky.

4 Metodológia

Prvým krokom v našej metodológii je zostavenie rozsiahleho datasetu obsahujúceho reálne a deep fake videá. Tento dataset musí byť dostatočne diverzifikovaný, aby zahŕňal rôzne osoby, emocionálne stavy a situácie. Videá budú získané z verejne dostupných zdrojov, ako sú video platformy, sociálne siete a databázy pre výskumné účely.

Dataset bude rozdelený na tréningovú, validačnú a testovaciu sadu. Tréningová sada bude použitá na trénovanie modelov, validačná sada na ladenie hyperparametrov a testovacia sada na vyhodnotenie výkonnosti finálneho modelu. Aby sme zabezpečili vysokú kvalitu a relevantnosť dát, každé video bude manuálne prekontrolované a označené expertmi na mikroexpresie.

Na detekciu mikroexpresií použijeme konvolučné neurónové siete (CNN), ktoré sú schopné analyzovať obrazové dáta a identifikovať jemné detaily vo výraze tváre. CNN budú trénované na označených datasetoch obsahujúcich mikroexpresie, pričom sa budú učiť rozpoznávať charakteristické vzory týchto jemných zmien.

Architektúra CNN bude optimalizovaná na detekciu rýchlych a krátkodobých zmien vo výraze tváre. Použijeme rôzne techniky, ako sú augmentácia dát a regulácia, aby sme zvýšili robustnosť a generalizáciu modelu. Výstupy modelu budú mikroexpresie klasifikované podľa ich typu a trvania, ktoré budú následne použité na ďalšiu analýzu.

Po detekcii mikroexpresií budeme analyzovať rozdiely medzi reálnymi a deep fake videami. Tento krok zahŕňa štatistickú analýzu frekvencie, trvania a typu mikroexpresií v oboch kategóriách videí. Použijeme rôzne štatistické metódy, ako sú t-testy, ANOVA a regresné analýzy, aby sme identifikovali signifikantné rozdiely.

Výsledky tejto analýzy budú vizualizované pomocou grafov a diagramov, ktoré nám umožnia lepšie pochopiť a interpretovať zistené rozdiely. Predpokladáme, že deep fake videá budú vykazovať nižšiu frekvenciu a kratšie trvanie mikroexpresií, ako aj vyššiu mieru nepresností a artefaktov.

Na základe zistených rozdielov vyvinieme algoritmus, ktorý bude schopný efektívne detekovať deep fake videá. Tento algoritmus bude kombinovať techniky strojového učenia a štatistickej analýzy na identifikáciu syntetických videí.

Algoritmus bude trénovaný na tréningových dátach a jeho výkon bude testovaný na testovacej sade. Budeme používať rôzne metriky, ako sú presnosť, precíznosť, recall a F1-score, aby sme vyhodnotili jeho výkonnosť. Naším cieľom je vytvoriť robustný a spoľahlivý nástroj na detekciu falošných videí, ktorý bude schopný pracovať v reálnom čase a poskytovať vysoko presné výsledky.

5 Experimenty a výsledky

Naše experimentálne nastavenie zahŕňa použitie výkonného hardvéru a softvéru na spracovanie a analýzu veľkých množstiev videodát. Používame výkonné grafické karty (GPU)

na urýchlenie tréningu konvolučných neurónových sietí a ďalšie výpočtovo náročné úlohy. Softvérové prostredie zahŕňa populárne nástroje a knižnice pre strojové učenie, ako sú TensorFlow a PyTorch.

V prvom kroku predspracúvame videodáta, ktoré zahŕňajú segmentáciu videí na jednotlivé snímky, normalizáciu obrazu a označenie mikroexpresií. Tieto predspracované dáta budú následne použité na tréning a testovanie našich modelov.

Po trénovaní konvolučných neurónových sietí na detekciu mikroexpresií predpokladáme dosiahnutie vysokej presnosti detekcie. Naša metóda by teda mala byť schopná spoľahlivo identifikovať mikroexpresie v reálnych videách s dostatočne vysokou presnosťou. Model by mal byť schopný klasifikovať rôzne typy mikroexpresií a analyzovať ich trvanie a frekvenciu.

Predpokladáme, že štatistická analýza rozdielov medzi reálnymi a deep fake videami ukáže signifikantné rozdiely vo frekvencii a trvaní mikroexpresií. Deep fake videá by mali vykazovať nižšiu frekvenciu mikroexpresií a kratšie trvanie týchto výrazov. Okrem toho by mali vykazovať vyššiu mieru nepresností a artefaktov v deep fake videách, čo by potvrdilo naše predpoklady o ťažkostiach reprodukcie mikroexpresií pomocou deep fake technológií.

Vyvinutý detekčný algoritmus testovaný na nezávislej testovacej sade by mal dosiahnuť vysokú presnosť a spoľahlivosť, pričom by mal vykazovať nízku mieru falošných pozitív a negatív.

6 Diskusia

Výsledky nášho doterajšieho výskumu naznačujú, že mikroexpresie môžu byť spoľahlivým indikátorom autenticity videí. Zistili sme, že deep fake videá majú výrazné ťažkosti s reprodukciou týchto jemných a rýchlych výrazov tváre, čo vedie k ich zníženej frekvencii a kratšiemu trvaniu. Tieto zistenia podporujú teóriu, že analýza mikroexpresií môže odhaliť syntetické videá, ktoré by inak mohli oklamať bežné detekčné metódy.

Naša metóda analýzy mikroexpresií poskytuje niekoľko výhod oproti existujúcim metódam detekcie deep fake videí. Vizuálna a zvuková analýza môže byť účinná, ale často zápasí s vysokou mierou falošných pozitív a negatív. Strojové učenie je veľmi silné, ale vyžaduje veľké množstvo dát na tréning. Naša metóda kombinuje výhody týchto prístupov a zameriava sa na jemné detaily, ktoré sú ťažko simulovateľné, čím poskytuje robustnejší a presnejší nástroj na detekciu.

Napriek doterajším pozitívnym výsledkom naša štúdia má niekoľko obmedzení. Naša analýza je závislá na kvalite a diverzite datasetu. Ak by dataset nebol dostatočne reprezentatívny, mohlo by to ovplyvniť výsledky. Ďalej, náš algoritmus bude trénovaný a testovaný na určitých typoch videí a situácií, takže jeho výkon v iných kontextoch môže byť odlišný. Budúci výskum by mal zahŕňať širšie spektrum videí a emocionálnych stavov, aby sa zabezpečila robustnosť a generalizácia modelu.

7 Budúce smerovanie a odporúčania

Naša štúdia poskytne základ pre ďalší výskum, ale existuje niekoľko oblastí, kde by bolo možné metódy zlepšiť. Napríklad, pokročilejšie techniky strojového učenia a hlbokého učenia by mohli zvýšiť presnosť a spoľahlivosť detekcie. Použitie hybridných modelov, ktoré kombinujú viaceré detekčné techniky, by mohlo tiež zvýšiť robustnosť systému.

Okrem detekcie deep fake videí by naša metóda analýzy mikroexpresií mohla nájsť uplatnenie aj v iných oblastiach. V psychológii a behaviorálnej vede by analýza mikroexpresií mohla pomôcť lepšie pochopiť emocionálne reakcie ľudí. V oblasti bezpečnosti a forenznej analýzy by mohla byť použitá na identifikáciu falošných výpovedí a odhalenie podvodov.

Výskum v oblasti detekcie deep fake videí si vyžaduje spoluprácu odborníkov z rôznych disciplín. Psychológovia, odborníci na strojové učenie, počítačoví vedci a forenzní analytici by mali spolupracovať na vývoji a zdokonaľovaní metód detekcie. Interdisciplinárny prístup môže

potom priniesť nové a inovatívne riešenia, ktoré budú schopné čeliť tejto rýchlo sa vyvíjajúcej hrozbe.

8 Záver

Tento článok skúma možnosti využitia mikroexpresií na detekciu deep fake videí. Na základe našich experimentov sme zistili, že mikroexpresie sú spoľahlivým indikátorom autenticity videí, pretože deep fake technológie majú výrazné ťažkosti s ich reprodukciou.

Výskum analýzy mikroexpresií predstavuje nový a sľubný prístup k detekcii deep fake videí. Naše výsledky naznačujú, že táto metóda môže poskytnúť účinný nástroj na boj proti syntetickým videám a prispieť k ochrane verejnosti pred dezinformáciami a manipuláciou. Budúci výskum by mal pokračovať v zdokonaľovaní týchto metód a ich aplikácií v rôznych oblastiach.

Literatúra

1. Ekman, P. (2009). *Telling lies: Clues to deceit in the marketplace, politics, and marriage*. W. W. Norton & Company.
2. Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A technique for the measurement of facial movement*. Consulting Psychologists Press.
3. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative adversarial nets. In Advances in Neural Information Processing Systems* (pp. 2672-2680).
4. Korshunov, P., & Marcel, S. (2018). *DeepFakes: A new threat to face recognition? Assessment and detection.* arXiv preprint arXiv:1812.08685. https://doi.org/10.48550/arXiv.1812.08685.
5. Nguyen, T. T., Nguyen, C. M., Nguyen, D. T., Nguyen, D. T., & Nahavandi, S. (2019). *Deep learning for deep fakes creation and detection: A survey.* arXiv preprint arXiv:1909.11573. https://doi.org/10.48550/arXiv.1909.11573.
6. Tariq, S., Lee, S., Kim, H., Shin, Y., & Woo, S. S. (2018*). Detecting both machine and human created fake face images in the wild.* In Proceedings of the 2nd International Workshop on Multimedia Privacy and Security (pp. 81-87). https://doi.org/10.1145/3267357.3267367.