Otkrivanje krivotvorenja slika i njihova lokalizacija

Kešina, Spomenko

SVEUČILIŠTE SJEVER

SVEUČILIŠNI CENTAR VARAŽDIN

DIPLOMSKI RAD br. 139-MMD-2024

OTKRIVANJE KRIVOTVORENJA SLIKA I NJIHOVA LOKALIZACIJA

Spomenko Kešina

Varaždin, Rujan 2024.

SVEUČILIŠTE SJEVER

SVEUČILIŠNI CENTAR VARAŽDIN

Studij Multimedija

DIPLOMSKI RAD br. 139-MMD-2024

OTKRIVANJE KRIVOTVORENJA SLIKA I NJIHOVA LOKALIZACIJA

Student:
Spomenko Kešina,
mat.br. 001603720

Mentor:
izv. prof. dr. sc. Emil Dumić

Varaždin, Rujan 2024.

Sveučilište Sjever
Sveučilišni centar Varaždin
104. brigade 3, HR–42000 Varaždin

Prijava diplomskog rada

Definiranje teme diplomskog rada i povjerenstva

ODJEL Odjel za multimediju

STUDIJ diplomski sveučilišni studij Multimedija

PRISTUPNIK Kešina Spomenko | JMBAG 0016037208

DATUM 26.08.2024. | KOLEGIJ Računalni vid

NASLOV RADA Otkrivanje krivotvorenja slika i njihova lokalizacija

NASLOV RADA NA ENGL. JEZIKU Image forgery detection and localization

MENTOR Emil Dumić | ZVANJE izv.prof.dr.sc.

ČLANOVI POVJERENSTVA

1. doc. art. dr. sc. Mario Periša - predsjednik
2. izv. prof. art. dr. sc. Robert Geček - član
3. izv. prof. dr. sc. Emil Dumić - mentor
4. doc. dr. sc. Andrija Bernik - zamjenski član
5.

Zadatak diplomskog rada

BROJ 139-MMD-2024

OPIS

U ovom radu će biti opisani i ispitani različiti modeli otkrivanja krivotvorenja slika i njihove lokalizacije, s naglaskom na modele dubokog učenja.

Digitalni sadržaj, kao što su slike, videozapisi i zvuk, svakodnevno se objavljuje na društvenim mrežama, gdje su slike najpopularniji dijeljeni resurs. Međutim, zahvaljujući ogromnoj popularnosti vrhunskih softvera za uređivanje slika, one se mogu lako uređivati bez ostavljanja vidljivih tragova. Stoga je često korisnicima teško ručno identificirati manipulirane slike. Krivotvorenje slika, manipulacija slikom ili slikovna forenzika je grana istraživanja u kojoj se manipulirane slike proučavaju za rješavanje izazovnih zadataka kao što su lokaliziranje promijenjenih regija, potvrđivanje autentičnosti slika te utvrđivanje izvora neovlaštenih slika. U radu će se opisati neke metode krivotvorenja slika poput kopiranja i pomicanja unutar iste slike, kopiranja dijelova iz različitih slika, uklanjanje objekata iz slika i popravljanje uklonjenih dijelova slika (inpainting), poboljšavanje slika i drugo. Opisat će se neke od postojećih baza krivotvorenih slika poput CASIA, Forensics, CoMoFoD, GRIP, COVERAGE i FAU baza slika. Potom će se opisati općeniti algoritmi dubokog učenja koji se mogu koristiti i kod otkrivanja krivotvorenja slika, kao što su metode bazirane na CNN, R-CNN, LSTM, autoenkoderu i drugo. Specifično, opisat će se novije metode poput TruFor, HiFi-IFDL, ImageForensicsOSN, PCL i CFL-Net. Opisat će se metode za usporedbu algoritama za lokalizaciju krivotvorenih dijelova slika: preciznost, odziv, F1 mjera i AUC-ROC krivulja za različite pragove detekcije.

U praktičnom dijelu rada će se koristiti neke od novijih metoda za otkrivanje i lokalizaciju krivotvorenih dijelova slika opisanih ranije. Koristit će se neka od postojećih baza ili će biti izrađena nova baza s nekom od metoda krivotvorenja slika. Rezultati algoritama će se usporediti koristeći preciznost, odziv, F1 mjere te AUC-ROC krivulju.

ZADATAK URUČEN 02.09.2024. | POTPIS MENTORA Emil Dumić

SVEUČILIŠTE SJEVER

Predgovor

Želim izraziti duboku zahvalnost profesoru Emilu Dumiću za njegovu neprocjenjivu podršku i vodstvo tijekom izrade ovog rada. Njegova stručnost, uvidi i ohrabrenje bili su ključni u oblikovanju smjera mog istraživanja i prevladavanju izazova s kojima sam se susreo. Iskreno sam zahvalan na njegovom mentorstvu i predanosti koji su značajno doprinijeli uspješnom dovršetku ovog rada.

Sažetak

U posljednjih nekoliko godina, manipulacija digitalnih slika postala je sveprisutna. Jednostavnost izmjene slika rezanjem, kloniranjem i mijenjanjem veličine brzo postaje izazov u provjeravanju njihovog integriteta i autentičnosti. Osim toga, forenzičari koriste digitalne slike u svojim istragama te ističu važnost forenzičke analize digitalnih slika u osiguravanju pouzdanosti digitalnih datoteka. Porast alata za uređivanje slika u posljednje vrijeme rezultirao je velikim brojem lažnih i izmijenjenih slika koje se šire putem različitih medijskih platformi i interneta. Tokom godina razvijeno je mnogo tehnika za određivanje autentičnosti slika i detektiranje manipuliranih područja. Duboko učenje se pokazalo učinkovitim u rješavanju širokog spektra kompleksnih problema, obuhvaćajući analizu velikih podataka, računalni vid i sofisticirane kontrolne sustave. Unatoč svojim napretcima, duboko učenje također je korišteno za razvoj softvera koji predstavlja rizike za privatnost, demokraciju i nacionalnu sigurnost. Ovdje se posebno ističe pojava tehnologije dubokih varki (engl. *DeepFake*). Ti algoritmi imaju sposobnost generiranja lažnih slika i videozapisa koji su gotovo vjerodostojni kao i stvarni. Zbog ovog razloga postoji hitna potreba za tehnologijama koje mogu automatski identificirati i provjeriti autentičnost digitalnog vizualnog sadržaja.

Ključne riječi: Aktivna forenzika, Detekcija, Duboka varka, Duboko učenje, Krivotvorenje, Lokalizacija, Manipulacija, Mreža, Pasivna forenzika, Vodeni žig, Zaštita

Summary

In recent years, digital image manipulation has become widespread. The ease of altering images through cropping, cloning, and resizing is rapidly becoming a challenge in verifying their integrity and authenticity. Additionally, forensic experts use digital images in their investigations, highlighting the importance of forensic analysis of digital images to ensure the reliability of digital files. The rise of image editing tools has recently led to a significant number of fake and altered images spreading across various media platforms and the internet. Over the years, many techniques have been developed to determine image authenticity and detect manipulated areas. Deep learning has proven effective in solving a wide range of complex problems, encompassing big data analysis, computer vision, and sophisticated control systems. Despite its advancements, deep learning has also been used to develop software that poses risks to privacy, democracy, and national security. A notable example is the emergence of deepfake technology. These algorithms can generate fake images and videos that are nearly as convincing as real ones. For this reason, there is an urgent need for technologies that can automatically identify and verify the authenticity of digital visual content.

Keywords: Active forensics, Detection, Deepfake, Deep learning, Forgery, Localization, Manipulation, Network, Passive forensics, Watermark, Protection

Popis korištenih kratica

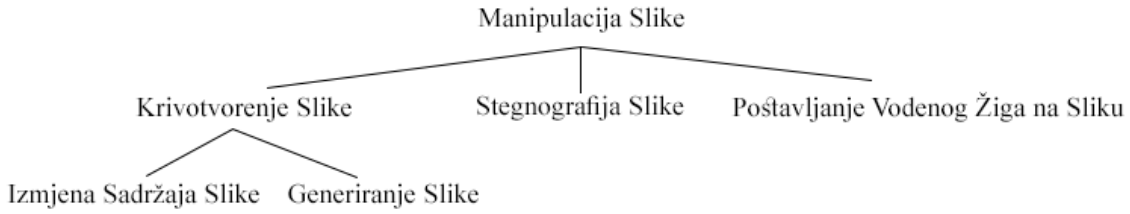
DIF Digital image forensics
Forenzička analiza digitalnih slika
GAN Generative adversrial network
Generativna suparnička mreža
CGI Computer-generated imagery
Računalno stvorene slike/videozapisi
GPU Graphics processing unit
Grafička procesorska jedinica
DCT Discrete cosine transform
Diskretna kosinusna transformacija
DWT Discrete wavelet transform
Diskretna valićna transformacija
PCET Polar complex exponential transform
Polarna kompleksna eksponencijalna transformacija
SIFT Scale-invariant feature transform
Transformaciju značajki neovisna o skali
SURF Speeded up robust features
Ubrzane robusne značajke
CFA Color filter array
Interpolacija filtera boja
AUC Area under curve
Površina ispod krivulje
ROC Receiver operating characteristic
Operativne karakteristike prijemnika
FPR False positive rate
Lažno pozitivan omjer
TPR True positive rate
Istinito pozitivan omjer
SCI Source camera identification
Identifikacija izvornog fotoaparata
PRNU Photo response non-uniformity
Neujednačenost odziva fotografije
CRF Camera response function
Funkcija odziva kamere
LPQ Local phase quantization
Lokalna kvantizacija faze
LPIP Local invariant planar irradiance points
Lokalna invarijanta točaka planarne iluminacije
SWT Stationary wavelet transform
Stacionarna valićna transformacija
SRM Spatial rich model
Prostorno bogat model
MSE Mean squared error
Srednje kvadratna pogreška
SSIM Structural similarity index measure
Indeks strukturne sličnosti
BER Bit error rate
Stopa pogrešaka bita
NCC Normalized cross-correlation
Normirana Pearsonova korelacija
PSNR Peak signal-to-noise ratio
Omjer vršnog signala i šuma
LSB Least significant bits
Najmanje značajni bitovi
RPS Random pixel selection
Nasumični odabir piksela
PMM Pixel mapping method
Tehnika mapiranja piksela
PVD Pixel value differencing
Razlikovanje vrijednosti piksela

GLM Gray level modification
Modifikacija sive razine
DFT Discrete Fourier transform
Diskretna Fourierova transformacija
DWT Discrete wavelet transform
Diskretna wavelet transformacija
DCvT Discrete curvelet transform
Diskretna curvelet transformacija
SS Spread spectrum
Prošireni spektar
SVD Singular values decomposition
Singularna dekompozicija matrice
SVM Support vector machines
Strojevi s potpornim vektorima
RF Random forests
Slučajnih šume
CNN Convolutional neural network
Konvolucijska neuronska mreža
LFR Laplacian filter residuals
Rezidualni Laplacianov filter
PCA Principal Component Analysis
Analizu glavnih komponenata
FAST Features from Accelerated Segment Test
Značajke ubrzanog segmentnog testa
HOG Histogram of Gradients
Histogram gradijenata
ORB Oriented FAST and rotated BRIEF
Orijentirani FAST i Rotirani BRIEF
BRIEF Binary robust independent elementary features
Robusne binarne nezavisne elementarne značajke
PRNU Photo-response non-uniformity
Parametri neujednačenog izlaza kamere
RANSAC Random Sample Consensus
Konsenzus slučajnih uzoraka
CMFD Copy-move forgery detection
Otkrivanje krivotvorenja kopiranje-pomicanjem
DCNN Deep convolutional neural network
Uuboka konvolucijska neuronska mreža
CNN Convolutional neural network
Konvolucijska neuronska mreža
ASPP Atrous spatial pyramid pooling
Prošireno prostorno piramidalno udruživanje
SEAM Squeeze and excitation attention modules
Moduli za pažnju kompresije i ekscitacije
FCN Fully convolutional networks
Potpuno konvolucijske mreže
OSN Online social networks
Društvene mreže
DNN Deep neural network
Duboka neuronska mreža
SSC Segmentation-based Self Calibration
Samokalibracija temeljena na segmentaciji
CSC Classification-based Self Calibration
Samokalibracija temeljena na klasifikaciji

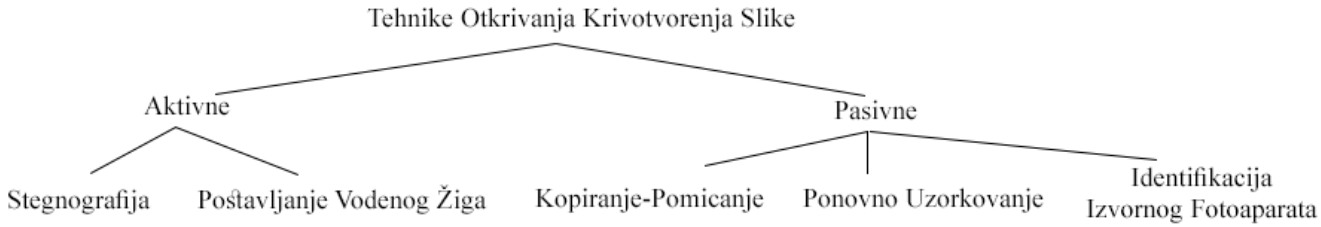
Sadržaj

1. Uvod

U današnjem tehnološki pismenom svijetu, digitalne kamere stekle su ogromnu popularnost te su integrirane u različite uređaje poput pametnih telefona, nadzornih kamera i drugih. Paralelno s ovim porastom sofisticirani softveri za uređivanje slika također su postali široko dostupni. Ova rasprostranjenost digitalnih slika otvorila je nove puteve poput pomoći forenzičarima u kriminalističkim istragama. Ipak, ručna analiza velikog broja digitalnih slika pokazuje se kao mukotrpna zadaća. Na primjer, tijekom tipičnog istraživačkog projekta koji traje od 6 do 18 mjeseci, forenzički stručnjaci mogu pregledati preko 100.000 digitalnih datoteka. Ilustrativna forenzička studija otkrila je bazu podataka koja sadrži impresivnih 300.000 slikovnih datoteka i 110 videozapisa pri čemu samo 148 slika prikazuje nezakoniti sadržaj (seksualno zlostavljanje). Nadalje, forenzički zaključci mogu biti pod utjecajem različitih faktora poput kognitivnih pristranosti, obuke i organizacijskih postavki. U ovom području forenzička analiza digitalnih slika (engl. *digital image forensics*, DIF) pojavljuje se kao specijalizirano područje posvećeno pronalaženju i analiziranju digitalnih dokaza u kriminalističkim istragama. DIF se prvenstveno koncentrira na dva ključna izazova: autentificiranje izvora slike i osiguravanje njezinog integriteta. Otkrivanje izvora digitalne slike uključuje identifikaciju jedinstvenih karakteristika poput modela kamere korištene za snimanje. S druge strane, procjena integriteta digitalne slike uključuje pregled njezinih sadržaja radi otkrivanja bilo kakvih znakova manipulacije ili falsificiranja koji bi doveli do izmišljene slike. Pojava manipuliranih slika na društvenim mrežama i platformama za razmjenu poruka naglo je porasla, otvarajući put porastu lažnih vijesti. Manipulacija slikama obuhvaća različite operacije poput kopiranja i pomicanja ili ponovnog uzorkovanja koji mijenjaju digitalne slike. Hijerarhijska struktura za manipulaciju slikama prikazana na slici 1.1 prikazuje razlike između krivotvorenja slike gdje je cijela slika zlonamjerno izmijenjena i manipulacije slikama gdje su samo određeni dijelovi namjerno promijenjeni. Generiranje slika uključuje stvaranje slika pomoću računalnog softvera ili algoritama kako bi se imitirale scene iz stvarnog svijeta. U steganografiji se slike modificiraju kako bi se sakrile tajne informacije koje su neprimjetne ljudskom oku. Vodeni žig na slici uključuje dodavanje oznake radi provjere autentičnosti. Metode otkrivanja krivotvorenja slika kategoriziraju se u aktivne tehnike koje zahtijevaju prethodno znanje poput vodenih žigova i pasivne tehnike, a koje procjenjuju autentičnost bez prethodnih podataka. Ove metode identificiraju uobičajene operacije krivotvorenja poput kopiranja i pomicanja ili prostornih transformacija (promjene veličine, rotacije i rastezanja) kako je prikazano na slici 1.2. Duboko učenje je postalo moćno sredstvo za otkrivanje manipulacija slikama.



Slika 1.1: Hierarhijska struktura manipulacije slika, krivotvorenja i iskrivljavanja. (Izvor: [90].)



Slika 1.2: Tehnike otkrivanja krivotvorenih slika. (Izvor: [90].)

2. Primjene detekcije krivotvorenja slika

Metode detekcije krivotvorenja slika općenito se kategoriziraju kao aktivne ili pasivne. Ove metode ne samo da identificiraju krivotvorena područja unutar slike već mogu i procijeniti izvorni sadržaj. Aktivne tehnike korištene su za zaštitu vizualnog sadržaja, a oslanjaju se na tehnologije poput digitalnih potpisa i vodenih žigova. Iako digitalni potpisi autentificiraju niz bitova kriptografski, podložni su čak i manjim promjenama što ih čini više prikladnima za aplikacije poput zaštite autorskih prava nego za provjeru semantike slike. Kako bi se riješile ove ograničenosti razvijene su robusnije metode poput digitalnog vodenog žiga koji suptilno ugrađuje sigurnosne informacije unutar samog sadržaja kako bi otkrio manipulacije. Unatoč varijacijama poput robusnih potpisa i krhkog vodenog žiga ove metode i dalje se suočavaju s izazovima povezanim s ranjivošću i krhkošću metapodataka. Iako aktivne metode prednjače u pružanju dodatnih informacije za otkrivanje krivotvorina, zahtijevaju računalnu obradu neizmijenjene slike za generiranje vodenog žiga ili potpisa što obično zahtijeva specifični hardver ili softver u fazi akvizicije. Dekodiranje autentifikacijskih informacija zahtijeva pristup ključu kreatora ili detektoru vodenog žiga što predstavlja sigurnosni rizik ako dođe u ruke potencijalnih zlonamjernih korisnika. Iako pouzdano treće lice može potvrditi integritet slike, poteškoće u skalabilnosti ometaju njegovu izvedivost za svakodnevno dijeljenje slika na internetu. Nova komercijalna rješenja koja koriste tehnologiju lanca blokova (engl. *blockchain*) imaju cilj eliminirati potrebu za trećim licem, ali detalji o njihovom djelovanju su još uvijek ograničeni. Metode lanca blokova u kontekstu zaštite slika uključuju generiranje bloka za svaku sliku bez mijenjanja same slike. Unatoč potencijalu za učinkovito otkrivanje krivotvorina ove tehnike nisu široko prihvaćene. To se može pripisati inherentnoj potrebi za krhkom autentifikacijom zbog kriptografske osnove lanca blokova i preostalih poteškoća u skalabilnosti. Ipak, napredak se postiže na ovom području. Pasivne metode, poznate kao forenzika, ne zahtijevaju dodatne podatke vezane uz sliku. Njihov primarni cilj je autentificiranje slika analizom njihovih inherentnih karakteristika tražeći tragove određene obrade slika. Ovaj pristup konvencionalno se smatra najpraktičnijim, posebno kod često dijeljenih slika. Napadači mogu manipulirati slike različitim sredstvima uključujući dupliciranje objekata, kompresiju ili transformacije kako bi zamutili izvornu krivotvorinu. Ova praksa naziva se anti-forenzika. Razrješavanje niza napada na sliku, poznato kao filogenija slike, predstavlja složeniji izazov od razlikovanja autentičnih i krivotvorenih slika. Ovaj zadatak uključuje identifikaciju više vrsta napada i njihov kronološki redoslijed. Na primjer, ako napadač izvrši tri različite manipulacije te svaku primjeni samo jednom, zadatak postaje znatno složeniji.

3. Vrste krivotvorenja slika

Uobičajene krivotvorine i manipulacije koje se obično susreću unutar aplikacija o kojima je nedavno bilo riječi:

- **Kopiranje-pomicanje (engl. *Copy-move*)** - Krivotvorenje kopiranjem-pomicanjem uključuje dupliciranje jednog ili više područja unutar slike i njihovo umetanje na različita područja iste slike. Ova obmanjujuća tehnika često se koristi za prikrivanje informacija ili repliciranje objekata i osoba što na kraju namjereno iskrivljuje značenje originalne slike.

- **Spajanje (engl. *Splicing)*** - Spajanje je tehnika slična krivotvorenju kopiranjem-pomicanjem, ali uključuje izrezivanje i lijepljenje područja ili objekata iz jedne ili više različitih slika. Ova manipulacija može se koristiti za prikrivanje određenog sadržaja ili predstavljanje izmišljenog scenarija.

- **Duboka varka (engl. *Deepfake*)** - Duboka varka predstavlja specifičan oblik manipulacije u kojem se modeli dubokog učenja koriste za stvaranje sintetičkog sadržaja u slikama ili videozapisima. Pojam "duboko" ističe prijelaz iz ere prije dubokog učenja kada su stručnjaci ručno obavljali takve zadatke koristeći profesionalne alate za uređivanje do današnje ere, gdje duboki modeli poput generativnih suparničkih mreža (GAN - engl. *generative adversrial network*) automatiziraju proces. Česta upotreba duboke varke uključuje zamjenu lica jedne osobe licem druge osobe (često poznate osobe) preuzete iz druge slike ili videozapisa. Druga varijacija duboke varke uključuje prenošenje izraza lica s jedne osobe na drugu u drugoj slici ili videozapisu što se postiže pomoću generativnih metoda poput GAN-ova ili algoritama za spajanje koji poboljšavaju realizam rezultirajućeg lica. Iako su duboke varke uglavnom izrađene u svrhu zabave ili komedije, bilo je slučajeva gdje su zloupotrebljavane kako bi prikazale pojedince u lažnim scenarijima što je narušilo njihov ugled i izazvalo skandale. Duboke varke su često prisutne u videozapisima zbog njihovog većeg potencijalnog utjecaja, posebno kada se spoje s odgovarajućim audiozapisom. Alati poput FakeApp-a i faceswap-GAN-a olakšali su stvaranje uvjerljivih dubokih varki što je dovelo do širenja takvog sadržaja na Internetu. Iako su DeepFake-ovi za statične slike manje uobičajeni, oni ostaju značajni za otkrivanje prijevara jer nalikuju određenoj vrsti spajanja.

- **Računalno stvorene slike/videozapisi (CGI - engl. *computer-generated imagery*)** - Ova metoda uključuje stvaranje realističnih vizuala renderiranjem računalno generiranih 3D scena s fokusom na postizanje fotorealizma. Nedavni napretci u tehnologiji grafičkih procesorskih jedinica (GPU - engl. graphics processing unit) i industriji videoigara učinili su tehnike poput praćenja zrake svjetlosti (engl. ray tracing) dostupnijima što je rezultiralo dosad neviđenim razinama realizma. Grafički alati poput Unity-a i Unreal Engine-a razvijaju se i sada su široko dostupni po niskim cijenama što omogućujuje stvaranje visoko uvjerljivih renderiranih slika i videozapisa. Ovi generirani vizuali mogu podsjećati na stvarne fotografije i predstavljati rizik jer ih zlonamjerni pojedinci mogu zloupotrijebiti kako bi izmislili lažne scenarije. Bitno je napomenuti da je i dalje potrebno stručno znanje za postizanje uvjerljivih rezultata s CGI sadržajem, za razliku od spajanja, budući da se ove scene potpuno grade iz ničega.

- **Stvaranje lica pomoću GAN-a** – Ovo je široko korištena metoda za generiranje lažnog sadržaja koja uključuje stvaranje autentičnih lica nepostojećih pojedinaca pomoću GAN mreža. Ova tehnika uključuje unošenje slučajnog vektora šuma (engl. *Noise vector*) u trenirani model koji zatim generira jedinstveno i realistično lice. Za razliku od CGI generiranog sadržaja ovaj proces generira nove lažne slike umjesto dupliciranja postojećih. Nvidia je predstavila revolucionarnu arhitekturu GAN-a za tu svrhu što je rezultiralo impresivnim rezultatima. Interaktivne demonstracije koje prikazuju ovu inovaciju mogu se pronaći online. Iako se povremeno mogu pojaviti neki artefakti u pozadini, generirana lica izuzetno su uvjerljiva i teško ih je razlikovati od stvarnih golim okom.

4. Forenzičke tehnike slike - Pasivna forenzika

Pasivne metode detekcije krivotvorenja slika mogu se kategorizirati u dvije glavne vrste: tradicionalne tehnike ručnog odabira značajki i pristupi temeljeni na dubokom učenju. Tradicionalni algoritmi za otkrivanje krivotvorenja kopiranjem-pomicanjem uključuju diskretnu kosinusnu transformaciju (DCT - engl. *Discrete cosine transform*) [1], diskretnu valićnu transformaciju (DWT – engl. *Discrete wavelet transform*) [2], polarnu kompleksnu eksponencijalnu transformaciju (PCET - engl. *Polar complex exponential transform*) [3], transformaciju značajki neovisnu o skali (SIFT – engl. *Scale-invariant feature transform*) [4] i ubrzane robustne značajke (SURF - engl. *Speeded up robust features*) [5]. Za otkrivanje krivotvorenja spajanjem konvencionalni algoritmi uključuju detekciju inkonzistentnosti putem interpolacije filtera boja (CFA - engl. *Color filter array*) [6] i analizu značajki šuma. Ove tradicionalne metode često zaostaju zbog ograničene generalizacije, slabe robusnosti i nedovoljne točnosti detekcije. Nasuprot tome algoritmi za detekciju temeljeni na dubokom učenju koriste autonomne sposobnosti učenja kako bi riješili ove nedostatke konvencionalnih tehnika. Napredak metoda dubokog učenja doveo je do značajnog napretka u područjima poput klasifikacije i segmentacije slika. Duboko učenje sada se češće koristi u otkrivanju krivotvorenja slika. Prvo se uspostavlja mreža za detekciju manipulacija koja obavlja zadatke poput ekstrakcije značajki, klasifikacije i lokalizacije. Treniranjem mreže s velikim skupovima podataka parametri mreže se uče, a njezine težine se spremaju kako bi se postigli optimalni rezultati. Slika koja se ispituje dovodi se u mrežu za detekciju manipulacija koristeći trenirani model mreže. Zadatak otkrivanja krivotvorenja može se promatrati kao izazov binarne klasifikacije na razini piksela, određujući je li piksel promijenjen ili ne. Kao rezultat toga kriteriji procjene za sustav detekcije krivotvorenja trebali bi uključivati mjere koje uzimaju u obzir različite klasifikacije uzoraka poput istinito pozitivnih, lažno pozitivnih, istinito negativnih i lažno negativnih.

4.1. Osnovni okvir za otkrivanje krivotvorenja slika

Tehnike dubokog učenja donijele su značajan napredak u raznim područjima, a posebno u klasifikaciji i segmentaciji slika. Jedna od novih primjena dubokog učenja je otkrivanje krivotvorenja slika. Slika 4.1 prikazuje osnovnu strukturu za otkrivanje krivotvorenja slika temeljenu na dubokom učenju. Prvi korak uključuje izgradnju mreže za otkrivanje krivotvorenja koja provodi ekstrakciju značajki, klasifikaciju i lokalizaciju putem modela mreže. Treniranjem na velikim skupovima podataka mreža uči parametre i pohranjuje težine kako bi postigla optimalne rezultate. Ulazna slika se zatim unosi u mrežu, a pohranjeni model se koristi za otkrivanje bilo kakvog krivotvorenja.

Slika 4.1: Osnovni okvir za otkrivanje krivotvorenja na slikama temeljen na dubokom učenju. (Izvor: [91])

4.2. Mjerni podaci za evaluaciju performansi

Zadatak otkrivanja krivotvorina može se promatrati kao problem binarne klasifikacije gdje se svaki piksel klasificira kao manipuliran ili autentičan. Kao rezultat, mjere koje se koriste za evaluaciju učinkovitosti algoritama za otkrivanje krivotvorina moraju uzeti u obzir broj točno i netočno klasificiranih uzoraka uključujući istinite pozitivne (engl. *True positives*), lažno pozitivne (engl. *False positives*), istinite negativne (engl. *True negatives*) i lažno negativne (engl. *False negatives*). Najčešće korištene mjere za otkrivanje krivotvorina su preciznost (p – engl. *Precision*), odziv (r – engl. *Recall*) i F_1 rezultat (izraženi kao jednadžba (4.1)-(4.4)) koji pružaju mjere preciznosti, osjetljivosti i njihove kombinacije.

$$p = \frac{T_P}{T_P + F_P} \qquad (4.1)$$

$$r = \frac{T_P}{T_P + F_N} \qquad (4.2)$$

$$F_{pr} = \frac{F_P}{(T_N + F_P)} \qquad (4.3)$$

$$F_1 = \frac{2pr}{p + r} \qquad (4.4)$$

Broj manipuliranih piksela koji su otkriveni kao manipulirani označen je kao T_P (true positive), dok je broj autentičnih piksela koji su otkriveni kao manipulirani označen kao F_P (false positive).

Broj manipuliranih piksela koji su otkriveni kao autentični označen je kao F_N (false negative). Vrijednosti za p, r i F_1 kreću se u rasponu od 0 do 1. Viša vrijednost za p, r i F_1 ukazuje na bolji rezultat. Površina ispod krivulje (AUC – engl. *Area under curve*) važna je mjera koja procjenjuje učinkovitost binarnog klasifikatora. Odnosi se na područje ispod krivulje operativnih karakteristika prijemnika (ROC – engl. *Receiver operating characteristic*) i može učinkovito predstavljati točnost klasifikacije. Slično F_1 rezultatu, ROC krivulja uzima u obzir lažno pozitivan omjer (engl. *false positive rate*, FPR) i istinito pozitivan omjer tj. odziv (engl. *true positive rate*, TPR). Vrijednosti AUC-ROC-a se kreću od 0,5 do 1 pri čemu viša vrijednost ukazuje na bolje performanse. Vrijednost od 0,5 označava slučajno pogađanje i nema praktičnog značaja.

4.3. Skupovi podataka za otkrivanje krivotvorenja slika

Ovaj odjeljak raspravlja o raznim skupovima podataka koji se koriste za otkrivanje krivotvorina. Ovi skupovi podataka uključuju mješavinu originalnih i krivotvorenih slika kao i binarne oznake i djelomično uređene slike. Ovisno o specifičnom problemu koji se rješava koriste se različiti skupovi podataka. Za procjenu učinkovitosti algoritama za otkrivanje krivotvorina koriste se razni javno dostupni skupovi podataka krivotvorenih slika kako bi se testirala njihova izvedba. Tablica 4.1 pruža informacije o 14 skupova podataka, uključujući vrstu i količinu krivotvorina, format slike i rezoluciju.

Tablica 4.1: Skupovi podataka za otkrivanje krivotvorenja slika.

Skup podataka	**Godina**	**Tip krivotvorenja**	**Broj Krivotvorenih Slika/Autentičnih Slika**	**Rezolucija slike**	**Format Slike**
Columbia color [96]	2006	Spajanje	183/180	757 × 568–1152 × 768	BMP, TIF
MICC-F220 [99]	2011	Kopiranje-pomicanjem	110/1100	480 × 722–1070 × 800	JPG
MICC-F600 [99]	2011	Kopiranje-pomicanjem	160/440	722 × 480–800 × 600	JPG, PNG
MICC-F2000 [99]	2011	Kopiranje-pomicanjem	700/1300	2048 × 1536	JPG
CASIA V1 [95]	2013	Kopiranje-pomicanjem, Spajanje	921/800	284 × 256	JPG
CASIA V2 [95]	2013	Kopiranje-pomicanjem, Spajanje	5123/7200	320 × 240–800 × 600	JPG, BMP, TIF
Carvalho [100]	2013	Spajanje	100/100	2048 × 1536	PNG
CoMoFoD [101]	2013	Kopiranje-pomicanjem	4800/4800	512 × 512–3000 × 2500	PNG, JPG
COVERAGE [97]	2016	Kopiranje-pomicanjem	100/100	2048 × 1536	TIF
Korus [102]	2017	Kopiranje-pomicanjem, Spajanje	220/220	1920 × 1080	TIF
USCISI [103]	2018	Kopiranje-pomicanjem	100,000/-	320 × 240–640 × 575	PNG
MFC 18 [104]	2019	Višestruka manipulacija	3265/14,156	128 × 104–7952 × 5304	RAW, PNG, BMP, JPG, TIF
DEFACTO [105]	2019	Višestruka manipulacija	229,000/-	240 × 320–640 × 6405	TIF
IMD 2020 [106]	2020	Višestruka manipulacija	37,010/37,010	193 × 260–4437 × 2958	PNG, JPG

4.4. Pasivne tehnike

Pasivne metode autentifikacije razlikuju se od aktivnih metoda po tome što se ne oslanjaju na nikakve informacije iz originalne slike. Umjesto toga koriste inherentne karakteristike slike kako bi se analizirao njen sadržaj bez potrebe za unaprijed umetnutim markerima. Sljedeći odlomci detaljno će opisati značajne pasivne tehnike, uključujući identifikaciju izvornog fotoaparata, detekciju kopiranja i premještanja, te detekciju ponovnog uzorkovanja.

4.4.1. Identifikacije izvornog fotoaparata

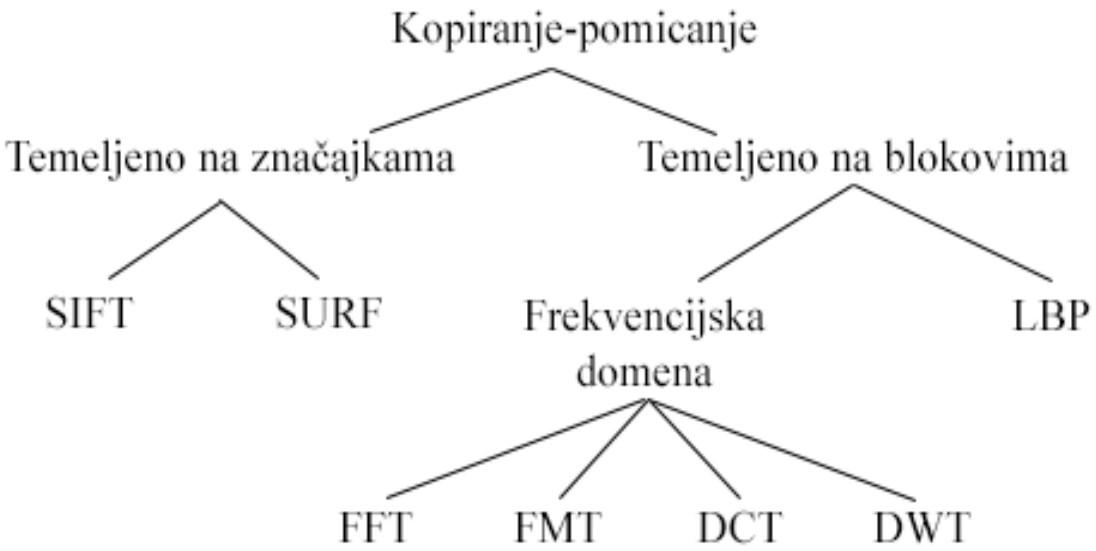
Tehnike identifikacije izvornog fotoaparata (SCI – engl. *Source camera identification*) koriste se za prepoznavanje jedinstvenih karakteristika digitalnih slika koje ostavljaju odgovarajući fotoaparati koji su ih snimili. Ove metode prate dokaze ostavljene u svakoj fazi snimanja slike kako bi stvorile potpis koji može točno odrediti fotoaparat odgovoran za sliku. Tablica 4.2 daje pregled glavnih SCI tehnika i njihovih povezanih istraživanja. Među ovim tehnikama su neujednačenost odziva fotografije (PRNU – engl. *Photo response non-uniformity*), ekstrakcija značajki i funkcija odziva kamere (CRF – engl. *Camera response function*). Kako bi se suprotstavili korištenju PRNU-a kao sredstva za provjeru izvora kamere, Taspinar et al. [7] predložili su protumjeru. Dodatno, Xu et al. [8] koristili su značajke teksture izvučene iz obojenih slika, konkretno lokalne binarne uzorke (LBP - engl. *Local binary patterns*) i lokalnu kvantizaciju faze (LPQ – engl. *Local phase quantization*) u HSV prostoru boja za generiranje potpisa kamere. Hsu i Chang [9] razvili su pristup koji procjenjuje CRF iz lokalnih invarijantnih točaka planarne iluminacije (LPIP – engl. *Local invariant planar irradiance points*), koje se dobivaju segmentiranjem digitalne slike u različite regije. Iz tih regija CRF se izračunava pomoću LPIP-ova.

Tablica 4.2: Istraživanja koja se odnose na identifikaciju izvornog fotoaparata.

Tehnika	**Kratki opis**	**Referenca**
PRNU	Ovaj proces koristi šumove koje generira senzor fotoaparata tijekom procesa akvizicije kako bi proizveo potpis koji identificira fotoaparat.	[7]
Ekstrakcija značajki	Koristi skup značajki ekstraktiranih iz slike za pronalaženje potpisa na slikama koje su izrađene pomoću istog senzora.	[8]
CRF	Svjetlosni podražaji su zabilježeni senzorima u nelinearnoj funkciji, a informacije koje se odnose na to mapiranje koriste se za izradu potpisa digitalne kamere.	[9]

4.4.2. Detekcija kopiranja-premještanja

Primarni fokus metode otkrivanja kopiranja i premještanja je identificirati sve dijelove slike koji su kopirani i zalijepljeni unutar iste slike. Ovi kopirani dijelovi mogu biti izmijenjeni ili ostavljeni netaknuti. Generalni cilj ovog pristupa je prikriti određena područja slike čime se narušava njena autentičnost. Nakon manipulacije kopiranjem i premještanjem mogu se koristiti dodatne tehnike uređivanja poput promjene veličine, filtriranja i dodavanja šuma kako bi se dodatno prikrili bilo kakvi dokazi o neovlaštenim promjenama. Kao što je prikazano na slici 4.2, postoje dvije glavne metode za otkrivanje operacija kopiranja i premještanja u slici: pristupi temeljeni na značajkama i pristupi temeljeni na blokovima.



Slika 4.2: Tehinka otkrivanja manipulacije kopiranje-pomicanje. (Izvor: [90].)

Različite tehnike koriste se za identifikaciju kopiranih i manipuliranih slika. To uključuje pristupe temeljene na značajkama koji koriste transformaciju značajki neovisnih o mjerilu (SIFT – engl. *Scale invariant feature transform*) ili druge metode koje su otporne na šum i geometrijske promjene. S druge strane detekcija temeljena na blokovima uključuje dijeljenje slike na preklapajuće ili nepreklapajuće blokove i analizu njihovih značajki u frekvencijskom području. Još jedna metoda je ekstrakcija LBP značajki iz blokova koji se također mogu koristiti za provjeru teksture blokova. Dodatne pojedinosti i istraživanja o identifikaciji krivotvorenja kopiranja i premještanja nalaze se u tablicama 4.3 i 4.4.

Tablica 4.3: Istraživanja temeljena na značajkama vezana za kopiranje-pomicanjm.

Tehnika	Kratki opis	referenca
SIFT	Radi s ključnim točkama izdvojenim iz slika koje su nepromjenjive u mjerilu te može pružiti zadovoljavajuće rezultate za varijacije rotacije i osvjetljenja.	[10]
SURF	Cilj ovog procesa je brzo izdvajanje lokalnih značajki slike koje su otporne na šum i geometrijske transformacije.	[11]

Tablica 4.4: Istraživanja temeljena na blokovima vezana za kopiranje-pomicanjm.

Tehnika	Kratki opis	referenca
Frekvencijska Domena	Koristi frekvencijsku domenu za prepoznavanje mogućih manipuliranih područja na slikama. Mogu se koristiti različite tehnike poput FFT, DCT i DWT.	[12]
LBP	Radi sa značajkama izdvojenim iz blokova slike u sivim tonovima kako bi se prepoznala krivotvorena područja na slikama.	[13]

Lokalne značajke slika se široko koriste za pretraživanje i prepoznavanje objekata zbog njihove otpornosti na različite geometrijske transformacije. Jedna popularna metoda, SIFT, je pristup temeljen na značajkama koji se često koristi za detekciju napada kopiranja i premještanja. Na temelju toga, Li i Zhou [10] predložili su tehniku temeljenu na ključnim točkama (engl. *Keypoint-based*) za otkrivanje i lokalizaciju krivotvorina kopiranja i premještanja koristeći novu hijerarhijsku metodu podudaranja točaka značajki koja smanjuje broj točaka radi poboljšanog podudaranja. SURF je još jedna metoda temeljena na značajkama, a koristi se za konstrukciju detektora kopiranja i premještanja. SURF značajke proizlaze iz ključnih točaka koje je izdvojio brzi Hesseov detektor koji aproksimira Hesseovu matricu. U usporednom istraživanju Jaseela i Nishadha [11] su evaluirali SIFT i SURF prema njihovoj učinkovitosti u otkrivanju krivotvorina kopiranja i premještanja. Rezultati pokazuju da su algoritmi temeljeni na SURF-u značajno brži od onih temeljenih na SIFT-u. Identifikacija dupliciranih područja u digitalnim slikama može se postići korištenjem frekvencijske domene i transformacije signala. Mahmood et al. [12] koristili su stacionarnu valićnu transformaciju (SWT – engl. *Stationary wavelet transform*) i diskretnu kosinusnu transformaciju (DCT) za otkrivanje i lokalizaciju kopiranja i premještanja. SWT je odabran zbog svoje sposobnosti rada u spektralnim i prostornim domenama dok su vektori značajki reducirani pomoću DCT-a. Tehnike redukcije dimenzija obično se koriste za poboljšanje učinkovitosti takvih metoda.

Operator LBP je široko korišten deskriptor teksture u sivim tonovima za slike na sivoj razini otporne na rotaciju. U studiji Farooq et al. [13] predložena je pasivna shema za otkrivanje krivotvorenja slika koja je kombinirala prostorno bogat model (SRM - engl. *Spatial rich model*) s teksturalnim značajkama temeljenim na LBP operatoru. Njihova istraživanja pokazuju da kombiniranje LBP-a s matricama supojavljivanja (engl. *Co-occurrence matrices*) značajno poboljšava sposobnost modela za otkrivanje različitih vrsta krivotvorenja s većom točnošću.

Važno je napomenuti da su određene metode koristile kombinaciju tehnika uključujući izdvajanje značajki (engl. Feature extraction) i metode temeljenih na blokovima, kako bi se riješio problem otkrivanja manipulacije kopiranje premještanjem.

4.4.3. Otkrivanje ponovnog uzorkovanja

Ponovno uzorkovanje se odnosi na primjenu prostornih transformacija kao što su promjena veličine, rotacija i rastezanje na digitalnoj slici, a može se identificirati tehnikama za otkrivanje ponovnog uzorkovanja. Te transformacije često ostavljaju tragove koji nisu prisutni u izvornoj slici te se mogu pojaviti u raznim scenarijima poput potrošačke elektronike, internetskih trgovina i servisa za dijeljenja slika. Kako bi identificirali ove tragove Peng et al. [14] predložili su korištenje višesmjernih visokopropusnih filtara i autoregresivnog modela za izdvajanje značajki i normiranih histograma. Međutim, njihov pristup je bio podložan JPEG kompresiji s gubitkom (engl. *Lossy JPEG compression*). Kao alternativu, Vazquez-Padin et al. [15] razvili su model temeljen na principima dekompozicije podprostora i teorije nasumičnih matrica koji je pokazao nisku računalnu složenost. Ipak, njihovi rezultati su također otkrili da upotreba nelinearnih operatora može umanjiti učinkovitost modela.

5. Forenzičke tehnike slike - Aktivna forenzika

Postavljanje digitalnog vodenog žiga (engl. *Digital watermarking*) ističe se kao izuzetno učinkovita metoda za aktivnu borbu protiv piratstva i zaštitu autorskih prava. Može se podijeliti u tradicionalne tehnike koje se oslanjaju na ručni odabir značajki i one temeljene na dubokom učenju. Inicijalno, postavljanje digitalnog vodenog žiga uglavnom je koristilo prostorno ugrađivanje poput najmanje značajnih bitova (LSB) [16], ali ovaj pristup bio je podložan otkrivanju putem analize parova uzoraka zbog nedostatka otpornosti [17]. Kako bi se poboljšala robusnost pojavile su se razne strategije temeljene na transformaciji domene, uključujući DWT [18], DCT [19], Contourlet transformaciju [20] i Hadamard transformaciju [21]. U novije vrijeme integracija tehnologije dubokog učenja značajno je unaprijedila postavljanje digitalnog vodenog žiga slika, donoseći značajne uspjehe.

5.1. Osnovni okvir robusnog algoritma za označavanje vodenim žigom

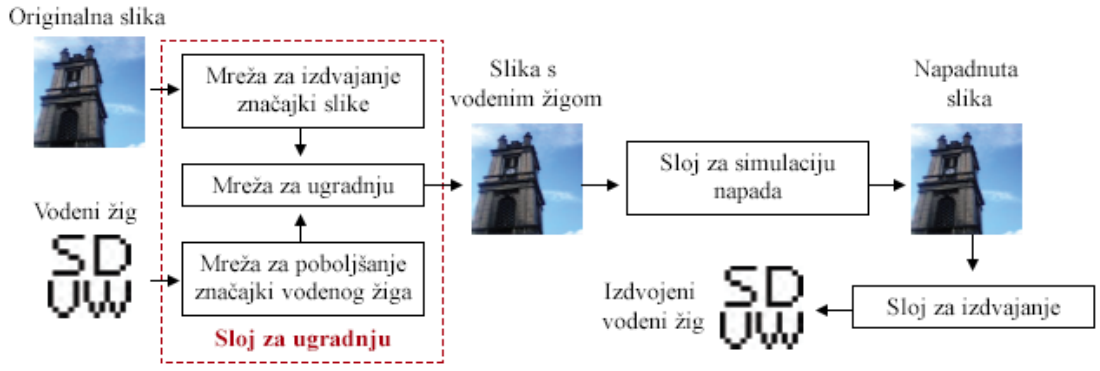
Osnovna struktura postavljanja robusnog vodenog žiga temeljenog na dubokom učenju uključuje tri ključne komponente (slika 5.1): sloj ugrađivanja (koji obuhvaća mrežu za izdvajanje značajki slike i mrežu za poboljšanje značajki vodenog žiga), sloj simulacije napada i sloj ekstrakcije. Ovaj model djeluje kroz dvije faze: propagaciju unaprijed i reverzno ažuriranje gradijenta tijekom iteracije parametara. U fazi propagacije unaprijed, originalna slika i slika s vodenim žigom podvrgavaju se obradi putem mreže za ekstrakciju značajki slike i mreže za poboljšanje značajki vodenog žiga kako bi se uhvatila visoka razina značajki slike i vodenog žiga. Nakon toga ovi izlazi unose se u mrežu za ugrađivanje vodenog žiga kako bi se proizvela slika s vodenim žigom. Sloj simulacije napada uvodi različite napade poput šuma, filtriranja, geometrijskih promjena i JPEG kompresije. Ovaj sloj generira suparničke uzorke dok je slika s vodenim žigom izložena tim napadima. Sloj ekstrakcije zatim izdvaja informacije vodenog žiga iz suparničkih uzoraka ili slika s vodenim žigom za autentifikaciju. Ovaj proces nastavlja trening modela propagacijom unatrag. U postupku propagacije unatrag (engl. *Backpropagation*), mjere poput gubitka srednje kvadratne pogreške (MSE - engl. *Mean squared error*) i indeksa strukturne sličnosti (SSIM – engl. *Structural similarity index measure*) koriste se za poboljšanje sličnosti između slika s vodenim žigom i originalnih slika fokusirajući se na gubitke u nijansama sive, kontrasta i strukture (izraženo u formulama (5.1), (5.2)). MSE gubitak koristi se za poboljšanje točnosti dohvaćanja vodenog žiga među izdvojenim i originalnim vodenim žigovima (formula (5.3)).

$$L_{MSE-I} = \sum_{i=0}^{P-1} \sum_{j=0}^{Q-1} [I(i,j) - I_W(i,j)]^2 \qquad (5.1)$$

$$L_{SSIM} = \frac{(2\mu_I \mu_{I_W} + C_1)(\sigma_{I,I_W} + C_2)}{(\mu_I^2 + \mu_{I_W}^2 + C_1)(\sigma_I^2 + \sigma_{I_W}^2 + C_2)} \qquad (5.2)$$

$$L_{MSE-W} = \sum_{i=0}^{L-1} [W(i,j) - W_e(i,j)]^2 \qquad (5.3)$$

gdje P označava širinu izvorne slike; Q označava duljinu izvorne slike; $I(i,j)$ i $I_W(i,j)$ označavaju svjetlinu na mjestu piksela (i,j) izvorne slike i slike s vodenim žigom; μ_I i μ_{Iw} označavaju srednju vrijednost sivih vrijednosti izvorne slike i slike s vodenim žigom, redom; σ_I^2 i σ_{Iw}^2 označavaju varijancu sivih vrijednosti izvorne slike i slike s vodenim žigom; σ_{I,I_w} označava kovarijancu izvorne slike i slike s vodenim žigom; C_1 i C_2 su konstante u rasponu $[10^{-4}, 9 \times 10^{-4}]$; L predstavlja duljinu vodenog žiga. Zatim je model izračunao odgovarajući gubitak točku po točku pomoću gradijenta od izlaznog kraja modela prema gore navedenom gubitku, a optimizator (obično Adam optimizator, SGD optimizator) je korišten za ažuriranje parametara modela kako bi se optimizirao zadatak modela (poboljšanje neprimjetnosti slike s vodenim žigom i točnosti ekstrakcije vodenog žiga nakon što je slika s vodenim žigom napadnuta).



Slike 5.1: Osnovni okvir za postavljanje robusnog vodenog žiga koji se temelji na dubokom učenju. (Izvor: [91])

Tri najvažnija pokazatelja evaluacije u algoritmu za digitalni vodeni žig slike su robusnost, neprimjetnost i kapacitet.

- **Robusnost** - Robusnost se koristi za mjerenje sposobnosti modela vodenog žiga da povrati originalni vodeni žig nakon što je slika podvrgnuta nizu namjernih ili nenamjernih procesiranja slike tijekom prijenosa putem elektroničkih ili neelektroničkih kanala. Stopa pogrešaka bita (BER – engl. *Bit error rate*) i normirana Pearsonova korelacija (NCC – engl. *Normalized cross-correlation*) obično se koriste kao objektivne evaluacijske mjere (izraženo u formulama (5.4) i (5.5)).

$$E_{BER}(\omega, \omega') = \frac{1}{L}\sum_{i=1}^{L}\left|\omega_i - \omega_i'\right| \qquad (5.4)$$

$$E_{NCC} = \frac{\sum_{i=1}^{L}(\omega_i - \bar{\omega})\left(\omega_i' - \bar{\omega}'\right)}{\sqrt{\sum_{i=1}^{L}(\omega_i - \bar{\omega})^2 \sum_{i=1}^{L}\left(\omega_i' - \bar{\omega}'\right)^2}} \qquad (5.5)$$

dje ω_i i $\bar{\omega}$ predstavljaju i -ti bit izvornog vodenog žiga i srednju vrijednost izvornog vodenog žiga; ω_i' i $\bar{\omega}'$ predstavljaju i-ti bit izdvojenog vodenog žiga i srednju vrijednost izdvojenog vodenog žiga; L predstavlja duljinu vodenog žiga.

- **Neprimjetnost** (engl. *Imperceptibility*) - Neprimjetnost se koristi za mjerenje senzornog utjecaja ugradbene točke na cijelu sliku nakon što je model završio ugrađivanje vodenog žiga (tj. garantirano je da se slika s vodenim žigom ne može razlikovati od originalne slike). Omjer vršnog signala i šuma (PSNR – engl. *Peak signal-to-noise ratio*) i indeksa strukturne sličnosti (SSIM) obično se koriste kao objektivne evaluacijske mjere (izraženo u formulama (5.6) i (5.7)).

$$E_{PSNR}(I, I_W) = 10log_{10}\frac{W \times H \times G_{MAX}^2}{\sum_{i=0}^{W-1}\sum_{j=0}^{h-1}[I(i,j) - I_W(i,j)]^2} \qquad (5.6)$$

$$E_{SSIM}(I, I_W) = \frac{\left(2\mu_I\mu_{I_W} + C_1\right)\left(2\sigma_{I,I_W} + C_1\right)}{\left(\mu_I^2 + \mu_{I_W}^2 + C_1\right)\left(\sigma_I^2 + \sigma_{I_W}^2 + C_2\right)} \qquad (5.7)$$

gdje W označava širinu izvornoe slike; H označava duljinu izvornoe slike; G_{MAX} označava maksimalnu sivu razinu izvorne slike; μ_I i μ_{I_w} označavaju srednju vrijednost sive vrijednosti izvorne slike i slike s vodenim žigom; σ_I^2 i $\sigma_{I_w}^2$ označavaju varijancu sive vrijednosti izvorne slike i slike s vodenim žigom; $I(i,j)$ I $I_W(i,j)$ označavaj svjetlinu na mjestu (i,j) izvorne slike i slike s vodenim žigom; σ_{I,I_w} označava kovarijancu izvorne slike i slike s vodenim žigom; C_1 i C_2 su konstante u rasponu $[10^{-4}, 9 \times 10^{-4}]$.

- **Kapacitet** - Kapacitet predstavlja maksimalni broj ugrađenih bitova modela vodenog žiga uz održavanje utvrđenih potrebnih mjera neprimjetnosti i robusnosti. Međusobno je ograničen s neprimjetnošću i robusnošću. Za povećanje robusnosti trebamo povećati kapacitet što smanjuje neprimjetnost. S druge strane, za povećanje neprimjetnosti, smanjujemo kapacitet i time smanjujemo robusnost. Broj ugrađenih bitova vodenog žiga po pikselu (bpp) obično se koristi za mjerenje kapaciteta modela (izraženo u formuli (5.8)).

$$E_{bpp} = \frac{W_{num}}{I_{num}} \qquad (5.8)$$

gdje W_{num} označava broj bitova vodenog žiga, a I_{num} označava ukupan broj piksela izvornog slike.

5.2. Aktivne tehnike

Kod aktivnih tehnika specifične komponente se ugrađuju u početnu sliku. To čini prethodno poznavanje slike ključnim za proces autentifikacije. Kao rezultat toga, aktivne tehnike obuhvaćaju integraciju vodenih žigova u sliku, bilo da su vidljivi ljudskom oku ili ne. Dva primjera aktivnih tehnika su steganografija i vodeni žig (engl. *Watermark*).

5.2.1. Steganografija

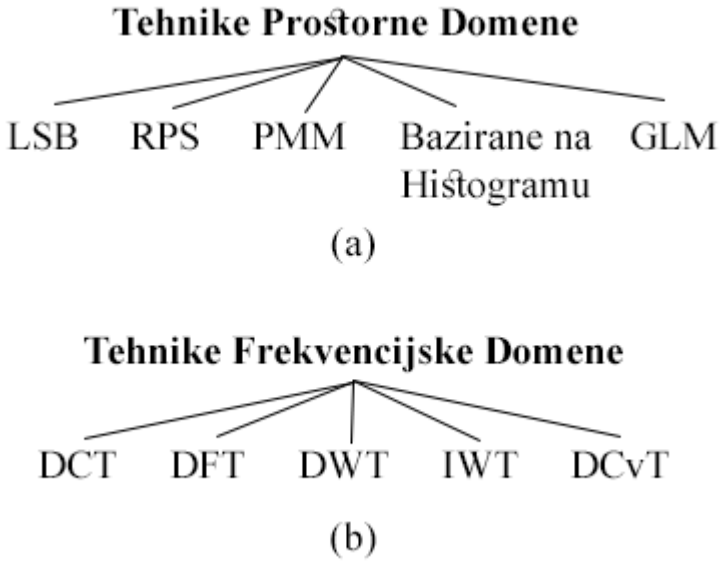
Steganografija je metoda prikrivene komunikacije koja koristi ograničenja ljudskog vizualnog sustava za šifriranje povjerljivih informacija. Za razliku od kriptografije koja nastoji učiniti poruke nečitljivima za neovlaštene strane, cilj steganografije je sakriti poruku na vidljivom mjestu.

To je čini korisnim alatom za održavanje povjerljivosti informacija u porukama kojima svatko može pristupiti. Usporedba steganografije i kriptografije prikazana je u Tablici 5.1.

Tablica 5.1: Komparativna analiza steganografije i kriptografije.

	Steganografija	**Kriptografija**
Sigurnosni mehanizam	Medij za prijenos ugrađuje tajnu poruku.	Obični tekst se pomoću tajnog ključa pretvara u šifrirani tekst.
Sigurnost podataka	Ovisi o tome kako je poruka ugrađena i koliko je poruka uočljiva.	Ovisi o robusnosti algoritma enkripcije i sofisticiranosti sigurnosnog ključa.
Robusnost	Ovisi o stupnju neprimjetnosti povjerljivih podataka.	Ovisi o složenosti algoritma enkripcije.
Napad	Manja podležnost napadima jer se podaci ne mogu otkriti.	Veća podležnost napadima jer šifrirani tekst može izazvati sumnju.

Steganografija uključuje skrivanje važnih podataka unutar određenog medija, nazvanog nositeljem, pomoću tajnog steganografskog ključa. Primarne steganografske metode prikazane su na slici 5.2 te kategorizirane kao tehnike prostorne (a) ili frekvencijske domene (b). Radovi usmjereni na steganografiju u prostornom području mogu se naći u Tablici 5.2, dok su članci koji se specifično fokusiraju na steganografiju u frekvencijskom području navedeni u Tablici 5.3.



Slika 5.2: Steganografske tehnike. (Izvor: [90].)

Tablica 5.2: Steganografske tehnike temeljene na prostornoj domeni.

Tehnika	**Kratki opis**	**Referenca**
LBS	Zamjenjuje manje značajne bitove piksela slike s bitovima informacija koje treba sakriti.	[22]
RPS	Manje značajni bitovi slike nasumično se odabiru i zamjenjuju bitovima koji sadrže informacije koje treba sakriti.	[23]
PMM	Odabire piksele slike za pohranu osjetljivih informacija na temelju matematičke funkcije koja omogućuje nedvosmisleno mapiranje tih piksela.	[24]
Bazirano na Histogramu	Ova tehnika ugrađuje tajne podatke unutar histograma slike.	[25]
GLM	Modificira sivu razinu kako bi se sakrile informacije i mapirali tajni podaci unutar slike.	[26]

Tablica 5.3: Steganografske tehnike temeljene na frekvencijskoj domeni.

Tehnika	**Kratki opis**	**Referenca**
DCT	Koristi informacije iz prikaza slike korištenjem zbroja sinusoida različitih magnituda i frekvencija za odabir piksela koji pohranjuju tajne informacije.	[27]
DFT	Ova metoda koristi informacije dobivene iz Fourierove transformacije za odabir piksela slike na koje će se pohraniti tajni podaci.	[28]
DWT	Pikseli koji pohranjuju tajne podatke odabiru se pomoću valovite dekompozicije slike.	[29]
IWT	Koristi informacije izdvojene iz valne transformacije cijelog broja za određivanje piksela slike koji će primiti osjetljive informacije.	[30]
DCvT	Koristi diskretnu transformaciju krivulje za određivanje piksela slike koji će pohraniti strateške informacije.	[31]

Tehnika steganografije koja koristi najmanje značajne bitove (LSB – engl. *Least significant bits*) slike često se koristi u prostornoj domeni. U svojem istraživanju, Carneiro Tavares i Madeiro Bernardino Junior [22] razvili su tehniku nazvanu LSB Word-Hunt koja je inspirirana igrom traženja riječi. Ova metoda učinkovito smanjuje broj modifikacija po pikselu što rezultira smanjenjem opterećenja sustava za 62,5% i poboljšanim performansama. LSB tehnika može se kombinirati s nasumičnim kodiranjem kako bi se omogućio nasumični odabir piksela (RPS – engl. *Random pixel selection*). Laskar i Hemachandran [23] predložili su RPS metodu koja skriva tajne informacije u najmanje značajnim bitovima crvene komponente u RGB slici. Za povećanje sigurnosti Bhattacharyya [24] istražio je korištenje tehnika mapiranja piksela (PMM – engl. *Pixel mapping method*) u različitim domenama koristeći različite mjere za sličnost slika.

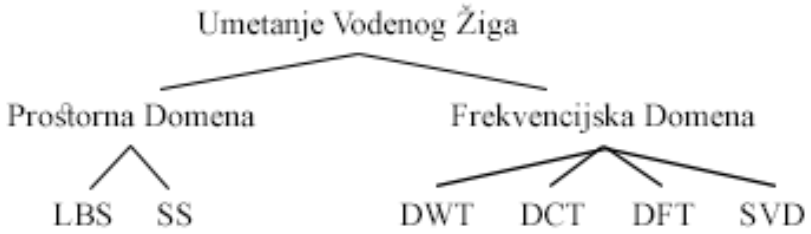
Tajni podaci se ubacuju samo u crvenu komponentu slike koristeći matematičku funkciju koja uzima u obzir intenzitet vrijednosti izvornog piksela (engl. *Seed pixel*) i njegovih susjednih piksela. Rezultati eksperimenta otkrili su da je predložena metoda imala veći kapacitet umetanja u usporedbi s drugim postojećim metodama kao što su razlikovanje vrijednosti piksela (PVD – engl. *Pixel value differencing*) i modifikacija sive razine (GLM – engl. *Gray level modification*). U PVD metodi razlika između dva susjedna piksela koristi se za određivanje maksimalnog broja bitova koji se mogu sakriti s tajnom informacijom. GLM metoda temeljena na modelu kojeg su predložili Shobana i Manikandan [26] koristi intenzitete RGB kanala za pohranu tajnih podataka, povećavajući kapacitet za skrivanje informacija. Qazanfari i Safabakhsh [25] također su predložili stegnoalitičku metodu baziranu na histogramima koja identificira i čuva osjetljive elemente u osnovnoj slici kako bi spriječila dodatno umetanje bitova čime se održava histogram osnovne slike. Al-dmour i Al-ani [30] također su predložili steganografski algoritam za slike, kombinirajući detekciju rubova i kodiranje logičkim operatorom XOR. Njihova metoda iskoristila je nižu osjetljivost ljudskog vizualnog sustava na promjene u područjima visokog kontrasta ugrađujući tajnu poruku u rubne piksele. Rezultati su pokazali značajna poboljšanja u kvaliteti slike i sigurnosti u usporedbi s drugim steganografskim metodama kao što su LSB i PVD.

Ni et al. [27] predstavili su steganografski pristup temeljen na frekvencijskoj domeni kombinirajući statistike prostornog i diskretnog kosinusnog transformacijskog područja (DCT). Ova metoda pretvara mjere izobličenje iz prostornog područja u DCT područje uključivanjem entropije blokovnog umetanja (engl. *Block embedding entropy*) iz različitih domena. Model također uzima u obzir statistike dekomprimiranih JPEG slika kako bi poboljšao sigurnost JPEG steganografije u prostornoj domeni. Ghoshal i Mandal [28] predložili su steganografsku tehniku za slike u boji temeljenu na frekvencijskoj domeni koristeći diskretnu Fourierovu transformaciju (DFT – engl. *Discrete Fourier transform*). Njihovi rezultati pokazali su bolje performanse u usporedbi s uobičajeno korištenom DCT metodom. Sličnim pristupom Ibaida i Khalil [29] koristili su diskretnu wavelet transformaciju (DWT – engl. *Discrete wavelet transform*) za razvoj metode koja kombinira tehnike šifriranja i miješanja kako bi osigurali povjerljive podatke pacijenata u elektrokardiogramskim signalima. Model je uspješno održavao sigurnost i povjerljivost komunikacije s manje od 1% izobličenja i dijagnosticiranim elektrokardiogramom nakon postavljanja vodenog žiga (engl. *Watermarking*). Jero i Ramu [31] uveli su drugačiji pristup koristeći diskretnu curvelet transformaciju (DCvT – engl. *Discrete curvelet transform*) za umetanje povjerljivih podataka pacijenata u elektrokardiogramske signale. DCvT dekomponira signal u frekvencijske podpojase (engl. *Frequency subbands*), a metoda kvantizacije umeće podatke pacijenata u koeficijente s vrijednostima oko nule u visokofrekvencijskom podpojasu.

Njihovi rezultati pokazali su superiornost ovog modela u usporedbi s metodama nasumičnog pozicioniranja.

5.2.2. Vodeni žig

Postavljanje vodenog žiga je metoda koja se koristi za autentifikaciju multimedijskog sadržaja. Ova metoda uključuje umetanje diskretne oznake kako bi se pomoglo u identifikaciji informacija bez izmjene primarnog sadržaja. Cilj je osigurati da se porijeklo sadržaja može lako provjeriti bez ikakvih sumnji. Urvoy et al. [32] opisuju četiri ključna elementa učinkovitih tehnika umetanja vodenih žigova: robusnost, sigurnost, kapacitet i nevidljivost. Slika 5.3 prikazuje različite tehnike watermarkinga u prostornim i frekvencijskim domenama. Tablice 5.4 i 5.5 prikazuju značajna djela u području watermarkinga.



Slika 5.3: Tehnike umetanja vodenog žiga. (Izvor: [90].)

Tablica 5.4: Tehnike umetanja vodenog žiga temeljene na prostornoj domeni.

Tehnika	**Kratki opis**	**Referenca**
LBS	Ova metoda koristi najmanje bitne bitove slike za pohranu informacija o autentičnosti.	[33]
SS	Tehnika spektralnog raspršenja koristi se za pohranjivanje informacija o izvoru slike unutar slike.	[34]

Tablica 5.5: Tehnike umetanja vodenog žiga temeljene na frekvencijskoj domeni.

Tehnika	**Kratki opis**	**Referenca**
DWT	Koristi valićnu transformaciju za pronalaženje područja interesa na slici koja mogu pohraniti vodeni žig.	[35]
DCT	Djeluje u domeni dobivenoj diskretnom kosinusnom transformacijom kako bi se postavio vodeni žig na sliku.	[36]
DFT	Koristi informacije izdvojene iz Fourierove transformacije za pronalaženje područja slike koje će pohraniti oznaku autentičnosti.	[32]
SVD	Koristi dekompoziciju pojedinačnih vrijednosti za pronalaženje područja slike na kojem će se nalaziti vodeni žig.	[37]

Tehnike postavljanja vodenog žiga koje koriste prostornu domenu obično koriste LSB i prošireni spektar (SS - engl. *Spread spectrum*). Bamatraf et al. [33] predložili su algoritam koji kombinira LSB i inverzni bit koji je evaluiran korištenjem vršnog omjera signal-šum (PSNR) i uspoređen s tradicionalnim LSB što je rezultiralo boljim performansama. Bose i Maity [34] predložili su model za detekciju vodenog žiga na degradiranim komprimiranim slikama, baziran na SS tehnici. Kada se razmatraju vodeni žigovi implementirani u frekvencijskoj domeni, četiri najčešće korištene tehnike uključuju DWT, DCT, DFT i singularnu dekompoziciju matrice (SVD – engl. *Singular values decomposition*). K.S. Bapat [35] proveo je komparativnu analizu watermarkinga koristeći DWT i DCT. Ernawan i Kabir [36] predstavili su okvir za watermarking baziran na DCT za zaštitu autorskih prava. Urvoy et al. [32] koristili su DFT za razvoj robusne metode umetanja vodenog žiga s informacijama sadržanim u faznoj komponenti. Khoo et al. [37] predložili su tehniku umetanja vodenog žiga baziranu na blokovima kombinirajući SVD s DWT. Sažetak komparativne analize između tehnika vodenog žiga i steganografije nalazi se u Tablici 5.6.

Tablica 5.6: Usporedna analiza vodenog žiga i steganografije.

	Vodeni žig	**Stegnografija**
Namjena	Zaštita autorskih prava i provjera autentičnosti.	Tajna komunikacija u javnom okruženju.
Nosioc	Vidljiva oznaka koja svima identificira izvor slike.	Skrivena tajna poruka koju može pročitati samo primatelj.
Komunikacija	Jedan prema mnogima.	Jedan na jedan.

6. Tradicionalne pasivne metode otkrivanja krivotvorenja

Tradicionalne pasivne metode koriste principe iz obrade signala, statistike, fizike i geometrije, često nazvane "klasičnim" ili "tradicionalnim" tehnikama. Potječući iz ere prije dubokog učenja, ove metode zahtijevaju minimalne podatke ili gotovo ništa za treniranje. Neke konvencionalne metode strojnog učenja poput klasteriranja, strojeva s potpornim vektorima (SVM – engl. *Support vector machines*), linearne/logističke regresije i slučajnih šuma (RF – engl. *Random forests*) još uvijek se oslanjaju na podatke za trening, ali se smatraju klasičnima zbog svojeg oslanjanja na modele s malim brojem parametara, stoga zahtijevaju manje podataka za obuku. Također, za razliku od modela dubokog učenja koji sami izdvajaju značajke, kod klasičnih modela strojnog učenja se značajke obično određuju ručno. Korisno je ukratko istaknuti ove tradicionalne pristupe iz dva glavna razloga:

1) Kao što je već rečeno, ove metode često zahtijevaju minimalne ili nikakve podatke za trening što je povoljno u scenarijima gdje je teško ili nemoguće dobiti dovoljan broj označenih slika za obuku složenog modela dubokog učenja. Nadalje, ovi pristupi općenito su manje računalno zahtjevni što ih čini prikladnima za implementaciju na uređajima manje snage kao što su pametni telefoni ili tableti.
2) Određeni osnovni koncepti i principi koji se koriste u ovim tehnikama također mogu nadopuniti modele dubokog učenja kako bi ubrzali proces učenja ili poboljšali ukupne performanse. Primjerice, SVM model služi kao krajnji korak faze klasifikacije koji se aplicira na izlazu konvolucijske neuronske mreže (CNN - engl. *Convolutional neural network*) [38]. Alternativno, pretvorba boja u YCbCr prostor boja i DCT transformacija djeluju kao predobrada prije CNN-a [39]. U drugom scenariju CNN obrađuje rezidualne Laplacianove filtere (LFR – engl. *Laplacian filter residuals*) izračunate iz ulaznih slika umjesto direktno samih slika.

U nadolazećim poglavljima pojedinačno je opisano pet primarnih kategorija koje su često povezane s pasivnim tradicionalnim metodama.

6.1. Metode temeljene na pikselima

Ove tehnike oslanjaju se na manipulaciju slika kako bi uvele anomalije koje mogu utjecati na statistička svojstva na razini piksela. Anomalije se mogu identificirati u prostornoj domeni, frekvencijskoj domeni ili kombinaciji obje. U slučajevima napada kopiranjem-pomicanjem često postoji primjetna korelacija između dupliciranih dijelova unutar slike. Međutim, zbog različitih veličina i oblika tih dijelova, postaje nepraktično i iscrpno istraživati sve moguće kombinacije

oblika i veličina. Metoda predložena od strane autora [40] u jednom istraživanju uključuje korištenje diskretne kosinusne transformacije (DCT). Ova tehnika uključuje podjelu slike na preklapajuće blokove, primjenu DCT-a na svaki blok i korištenje rezultirajućih koeficijenata kao vektora značajki za opis bloka. Duplicirane regije identificiraju se sortiranjem DCT blokovskih koeficijenata leksikografski i grupiranjem najsličnijih. Drugi pristup, kako je sugerirano u jednom istraživanju [41], koristi analizu glavnih komponenata (PCA – engl. *Principal Component Analysis*) na značajkama blokova slika, nakon čega uspoređuje reprezentacije blokova u smanjenom dimenzionalnom prostoru. Dok ove metode pokazuju otpornost na manje varijacije u dupliciranim regijama poput aditivnog šuma ili kompresije s gubicima, općenito su manje učinkovite kada su suočene s geometrijskim transformacijama poput rotacije ili skaliranja. Razmotrimo scenarij gdje je napad kopiranjem-pomicanjem poboljšan primjenom geometrijskih transformacija. Ove transformacije obično uključuju interpolaciju između susjednih piksela, često koristeći tehnike poput bilinearnog ili bikubnog interpoliranja. Ovisno o odabranoj metodi interpolacije pojavljuju se specifični korelacijski obrasci među tim pikselima. Statističke metode zatim se koriste za identifikaciju tih obrazaca i detekciju regija podvrgnutih geometrijskoj manipulaciji. Detekcija krivotvorenja temeljena na frekvenciji uključuje promatranje promijenjenih viših frekvencijskih koeficijenata u Fourierovoj analizi u spajanim regijama, pomažući u njihovoj detekciji unatoč potencijalno neprimjetnim granicama između spajanih i originalnih dijelova [42]. Metode temeljene na ključnim točkama još su jedan skup tehnika prilagođenih za otkrivanje napada kopiranjem-pomicanjem te obično uključuju niz specifičnih koraka:

1) Ekstrakcija ključnih točaka - Ključne točke često se opisuju kao značajne karakteristike unutar slike poput lokalnih minimuma ili maksimuma, kutova, oblika i ostalog. Neke popularne metode za ekstrakciju ključnih točaka uključuju široko korišteno skaliranje invarijantne transformacije značajki (SIFT) [43], ubrzane robustne značajke (SURF) [44] te značajke ubrzanog segmentnog testa (FAST – engl. *Features from Accelerated Segment Test*) [86].
2) Ekstrakcija deskriptora - Svaka ključna točka povezana je s jednim ili više vektora značajki (deskriptora) koji se izvlače. Ovi vektori obično pružaju sažet prikaz područja oko ključne točke. Osim vrijednosti značajki SIFT/SURF, drugi široko korišteni deskriptori uključuju histogram gradijenata (HOG – engl. *Histogram of Gradients*) i ORB (engl. *Oriented FAST and rotated BRIEF*) baziran na FAST-u i BRIEF-u (engl. *Binary robust independent elementary features*) [45].

3) Usklađivanje deskriptora – Tijekom ove faze deskriptori se procjenjuju na temelju mjere udaljenosti (ili odgovarajuće mjere sličnosti). Kada udaljenost između dva ili više deskriptora padne ispod određenog praga, utvrđuje se podudaranje između tih deskriptora.
4) Korak filtriranja - Tijekom ove faze filtracijski process se primjenjuje na rezultate podudaranja kako bi se eliminirala slabija podudaranja. Za tu svrhu koriste se različiti kriteriji poput Lowe-ovog omjera [43]. Lowe-ov omjer smatra podudaranje valjanim samo ako je udaljenost između dva najsličnija deskriptora značajno manja od udaljenosti između sljedećih najboljih deskriptora. Također se mogu koristiti dodatni kriteriji poput onih koji se fokusiraju na prostornu korelaciju između ključnih točaka.

6.2. Metode temeljene na formatu slike

Digitalni fotoaparati obično kodiraju slike u JPEG format tako što ih dijele u blokove od 8 × 8 piksela. Ti blokovi prolaze kroz DCT transformaciju i kvantizaciju rezultirajući izobličenjima duž granica susjednih blokova. Istraživači [46] su primijetili su da manipulacije poput kopiranja-pomicanja ili spajanja mijenjaju uzorak JPEG artefakata. Uveli su tehniku gdje se uzorak autentične slike koristi za procjenu JPEG kvantizacijske tablice. Zatim se slika segmentira u blokove, a mjerilo "izbličenja blokova" izračunava za svaki blok. Blok se označava kao manipuliran ako se njegova vrijednost značajno razlikuje od prosječne vrijednosti preko cijele slike. Međutim primjetna mana ovih pristupa je njihovo oslanjanje na pretpostavke specifične za format slike poput JPEG-a što ih čini manje univerzalno primjenjivima.

6.3. Metode temeljene na modelu kamere

Ove metode oslanjaju se na jedinstveni "otisak" ili "potpis" koji digitalni fotoaparati ostavljaju na svakoj generiranoj slici. Ova karakteristika može pomoći u povezivanju slike s određenim fotoaparatom koji ju je snimio. U jednom istraživanju [47], istraživači su koristili seriju slika iz poznatog fotoaparata kako bi predvidjeli parametre neujednačenog izlaza kamere (PRNU - engl. *Photo-response non-uniformity*). Ti parametri PRNU-a zatim se izvode iz ciljne slike, pretpostavljajući da potječu iz istog fotoaparata te se uspoređuju s prethodno izračunatim vrijednostima. Svaka razlika u procijenjenim parametrima ukazuje na potencijalno manipuliranje poput spajanja iz drugog tipa fotoaparata. Međutim, primjetno ograničenje ovog pristupa je njegova specifičnost za fotoaparat što zahtijeva zaseban skup podataka za trening za svaki tip fotoaparata kako bi se uspostavio njegov model PRNU-a. Osim toga ova se metodologija uglavnom koristi za otkrivanje napada spajanjem gdje je spajani dio fotografije preuzet s fotografije drugog modela fotoaparata, drugačijeg od onog koji se koristi za ciljanu sliku što

možda nije uvijek slučaj. U drugoj studiji [48], istraživači su iskoristili kromatsku aberaciju kako bi identificirali krivotvorene slike. Kromatska aberacija rezultat je nemogućnosti fotografskih leća da fokusiraju svjetlost različitih valnih duljina na istu točku senzora kamere. To dovodi do prostornih pomaka u RGB kanalima boja za svaku točku u sceni. Konstruiranjem modela koji približava učinke kromatske aberacije i procjenom njezinih parametara mogu se identificirati neusklađenosti u krivotvorenim područjima u odnosu na model. Slično kao i kod prethodne metode, ključno ograničenje je specifičnost za fotoaparat zbog varijacija u razinama kromatske aberacije među različitim kamerama na temelju njihovih leća. Stoga postavljanje preciznog praga za detekciju anomalija postaje izazovno ako fotoaparat koji je korišten za snimanje slike nije unaprijed poznat.

6.4. Metode temeljene na vrsti rasvjete

Uobičajeno je da u situacijama kada napadač izvede kopiranje-pomicanje ili spajanje, održavanje dosljednih uvjeta osvjetljenja između manipuliranog područja i okoline slike predstavlja značajan izazov. Čak i uz napredni softver poput Adobe Photoshopa ispravljanje ove razlike pokazalo se teškim. Kao rješenje, tehnike zasnovane na osvjetljenju (ili fizici) imaju za cilj uspostaviti sveobuhvatni model osvjetljenja iz glavne slike i identificirati bilo kakve lokalne razlike u odnosu na taj model kao indikacije manipulacije. Ponekad se integriraju metode poput konsenzusa slučajnih uzoraka (RANSAC – engl. *Random Sample Consensus*) [49] kako bi se poboljšala otpornost modela na izvanredne vrijednosti. Ove tehnike odlikuju širokom primjenjivošću jer se ne oslanjaju na pretpostavke o vrsti kamere koja je korištena za snimanje slike što omogućuje otkrivanje manipulacija poput kopiranja-pomicanja i spajanja. Nedostatak ovih pristupa leži u njihovoj ovisnosti o intrinzičnim fizičkim karakteristikama unutar slike. U scenarijima s kompliciranim uvjetima osvjetljenja (npr. u unutrašnjim prostorijama) gdje postavljanje globalnog modela osvjetljenja postaje neizvedivo, ove se metode ne mogu učinkovito primijeniti.

6.5. Metode temeljene na geometriji

Metode temeljene na geometriji iskorištavaju sklonost napada kopiranjem-pomicanjem ili spajanjem da unesu nepravilnosti u geometrijske karakteristike 3D scene snimljene na slici. U jednom istraživanju [50] autori su predstavili tehniku za određivanje glavne točke analizom poznatih ravnih objekata, primjećujući njezinu blizinu središtu slike. Pokazali su da pomicanje objekta u ravnini slike uzrokuje odgovarajuće prilagodbe glavne točke što može poslužiti kao potencijalni pokazatelj manipulacije.

Još jedna inovativna strategija [51] uključuje pretvaranje identificiranih objekata poput reklamnih ploča ili registracijskih pločica u ravnine korištenjem perspektivnih transformacija. Poravnavanjem tih referentnih objekata na odgovarajuću ravninu i kalibriranjem kamere postaje moguće dobiti mjerenja u stvarnom svijetu što pomaže u procjeni autentičnosti objekata unutar slike. Međutim, ove metode snažno se oslanjaju na pretpostavke o geometriji 3D scene i zahtijevaju ljudski unos za uspostavljanje mjera iz stvarnog svijeta za određene objekate na slici što dovodi do praktičnih ograničenja u njihovoj primjeni.

7. Metode temeljene na dubokom učenju

U posljednjem desetljeću tehnike dubokog učenja naglo su postale popularne i široko su korištene u različitim znanstvenim disciplinama. Njihova učinkovitost u rješavanju zadataka klasifikacije, regresije i segmentacije dobro je dokumentirana te često nadmašuje ljudsku točnost i preciznost. Ključna prednost modela dubokog učenja poput konvolucijskih neuronskih mreža (CNN) jest njihova sposobnost automatskog izdvajanja relevantnih značajki iz ulaznih podataka, eliminirajući potrebu da istraživači ručno odabiru značajke što je izazovan i intenzivan proces u tradicionalnim pristupima strojnom učenju. Ova sposobnost izdvajanja značajki osigurava da modeli uhvate bitne aspekte podataka prilagođene specifičnom zadatku koji se rješava te dodatno doprinose širokoj prihvaćenosti metodologija dubokog učenja. Novija literatura također istražuje primjenu tehnika dubokog učenja za detekciju krivotvorenih slika kako bi se povećala točnost iznad onoga što su tradicionalne metode nudile u prošlosti. Ove tehnike mogu se kategorizirati na temelju različitih kriterija, u ovom slučaju:

- **Vrsta detektiranog krivotvorenja:** kopiranje-pomicanje, spajanje ili oboje;
- **Svojstvo lokalizacije**, odnosno je li razmatrani algoritam sposoban lokalizirati područja koja su krivotvorena. U slučaju detekcije kopiranja-pomicanja dodatno pitanje je može li algoritam razlikovati između izvornog područja i ciljnog područja, tj. područja na koje je zalijepljen izvorni isječak. Ovo svojstvo je korisno primjerice u scenariju u kojem je forenzičkom stručnjaku zatraženo analiziranje manipulirane slike kako bi interpretirao semantičko značenje napada kopiranje-pomicanje;
- **Vrsta arhitekture**, odnosno algoritam koji je end-to-end rješenje te se može trenirati bez parametara koje treba ručno podešavati.

7.1. Otkrivanje krivotvorenja slika temeljeno na dubokom učenju

Pasivne tehnike detekcije krivotvorenih slika mogu se kategorizirati kao detekcija pojedinačnih krivotvorenja i detekcija generičkih krivotvorenja na temelju njihovih mogućnosti detekcije. Detekcija pojedinačnih krivotvorenja usredotočuje se na identifikaciju određenih vrsta manipuliranih slika poput krivotvorenja kopiranje-pomicanjem i spajanja slika. S druge strane, generičke metode detekcije krivotvorenja su svestrane i mogu otkriti različite vrste manipuliranih slika, uključujući kopiranje-pomicanjem, spajanje i uklanjanje.

7.1.1. Detekcija metode krivotvorenja kopiranje-pomicanjem

Metode otkrivanja krivotvorenja kopiranje-pomicanjem (CMFD - engl. *Copy-move forgery detection*) identificiraju manipulirane slike izdvajajući značajke povezane s manipulacijom. Izdvajanje značajki može se provesti globalno za cijelu sliku ili lokalno za određene regije. Odabir tehnika izdvajanja značajki značajno utječe na učinkovitost CMFD metoda. Te metode spadaju u dvije kategorije: tradicionalni ručni odabir značajki i metode temeljene na dubokom učenju. Tradicionalne metode temeljene na ručnom odabiru značajki mogu se dodatno klasificirati u metode temeljene na blokovima i metode temeljene na ključnim točkama. Iako metode temeljene na blokovima točno identificiraju manipulirane regije, suočavaju se s izazovima poput visoke računalne složenosti i poteškoća u rukovanju rotacijom i skaliranjem velikih razmjera. Kao odgovor uvedene su metode detekcije manipulacije temeljene na ključnim točkama. Ove metode koriste tehnike izdvajanja ključnih točaka kako bi se identificirale ključne točke i koriste algoritme za podudaranje značajki kako bi otkrile slične značajke. Iako metode temeljene na ključnim točkama odlično funkcioniraju u lociranju manipuliranih područja na uobičajenim slikama, suočavaju se s ograničenim brojem ključnih točaka na glatkim područjima što dovodi do nepronalaženja manipulacije i loše generalizacije algoritma. Predložene su CMFD metode detekcije manipulacije temeljene na dubokom učenju kako bi se riješila ograničenja tradicionalnih metoda temeljenih na ručnom odabiru značajki. S obzirom na to da se tehnike dubokog učenja nastavljaju brzo razvijati, pronašle su primjene u detekciji manipulacije unutar polja. Poboljšanja u performansama koja su pokazale CMFD metode detekcije manipulacija temeljene na dubokom učenju bila su izvanredna. Model temeljen na dubokom učenju ima sposobnost shvatiti temeljne značajke prisutne u slikama. Prepoznajući razlike između različitih vrsta slika mogu se identificirati manipulirane slike. U usporedbi s tradicionalnim pristupima, detekcija manipulacije CMFD temeljena na dubokom učenju nudi superiornu preciznost i detaljne reprezentacije značajki.

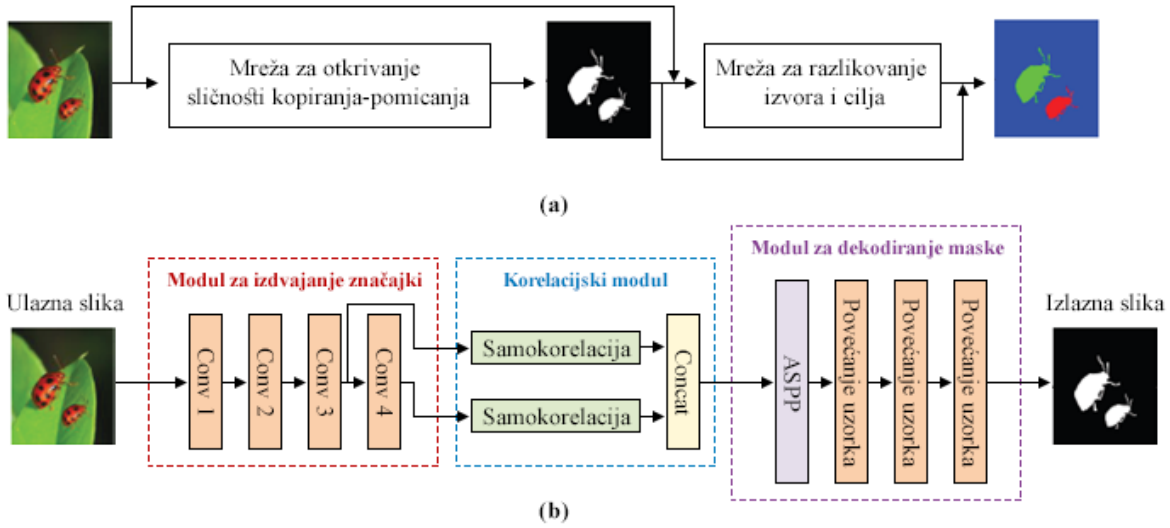
Da bi identificirali potencijalne manipulacije slika, Rao i Ni [52] su predstavili duboku konvolucijsku neuronsku mrežu (DCNN – engl. *Deep convolutional neural network*) model. Ovaj model koristi prostorne bogate modele (SRM – engl. *Spatial rich models*) s 30 osnovnih visokopropusnih filtera (engl. *High-pass filters*) za inicijalizaciju težina prvog sloja što efikasno smanjuje utjecaj semantike slike i ubrzava konvergenciju mreže. DCNN model izdvaja značajke slika kroz konvolucijski sloj i koristi potporni vektor stroj (SVM – engl. *support vector machine*) za klasifikaciju. Slično tome, Kumar i Gupta [53] su predložili model temeljen na konvolucijskoj neuronskoj mreži (CNN – engl. *Convolutional neural network*) za automatsko izdvajanje značajki u detekciji manipulacija slika s otpornošću na kompresiju slika, skaliranje, rotaciju i

zamućenje (engl. *Blurring*). Međutim, njihov pristup također zahtijeva analizu na razini piksela kako bi se identificirale manipulirane regije. Treba napomenuti da se oba rada [52, 53] fokusiraju isključivo na detekciju manipulacija bez mogućnosti lokalizacije manipuliranih regija što ograničava potencijal njihove primjene.

Li et al. [54] su predložili metodu koja kombinira segmentaciju slike i DCNN kako bi proširila primjenu algoritma i postigla lokalizaciju kopiranja i premještanja regija. Ovaj pristup koristi Super-BPD (engl. *Super boundary-to-pixel direction*) [55] za segmentaciju slike i dobivanje informacija o rubovima dok VGG-16 mreža i sloj mreže proširenog prostornog piramidalnog udruživanja (ASPP - engl. *Atrous spatial pyramid pooling*) [56] izdvajaju višeskalne značajke radi poboljšanja preciznosti. Uključivanje modula za podudaranje značajki (engl. *Feature matching module*) omogućuje preciznu lokalizaciju manipuliranih regija. Međutim, segmentacijski modul dodaje kompleksnost u računanju. Kako bi se tome suprotstavili Liu et al. [57] su razvili dvostupanjski okvir za detekciju koji koristi proširenu (engl. *dilated* ili *atrous*) konvoluciju i auto-korelacijsko podudaranje s prostornom pažnjom (engl. *Spatial attention*) u prvom stupnju. U drugom stupnju koristi se "superglue" metoda za eliminaciju lažnih upozorenja i popravak nepotpunih regija što dovodi do poboljšane točnosti. Ostali istraživači također su predložili nove metode za poboljšanje performansi detekcije i otpornosti algoritma. Zhong i Pun [58] uključili su višeskalnu konvoluciju u Inception arhitekturu dok su Kafali et al. [59] predstavili nelinearni Inception modul koristeći Volterra jezgru drugog reda za prepoznavanje linearnih i nelinearnih interakcija među pikselima. Nazir et al. [60] su poboljšali maskiranu regionalnu CNN, Mask R-CNN [61] (engl. *Mask region-based convolution network*) koristeći DenseNet-41 model za izdvajanje značajki. Zhong et al. [62] su razvili grubo-finu prostorno-kanalnu mrežu pažnje s modulom za pažnju, dizajniranim za pročišćavanje granica kako bi se poboljšalo otkrivanje finih krivotvorenih detalja. Iako su prethodne studije [52–54, 57–59] napredovale u poboljšanju performansi detekcije i otpornosti, nisu razlikovale između izvornih i manpuliranihih područja.

Predloženo je nekoliko metoda za precizno razlikovanje između izvornih i ciljanih područja. Wu et al. [63] su prvi uveli paralelnu mrežu za ovu svrhu. Njihov pristup uključuje granu za otkrivanje manipulacije koja identificira potencijalna manipulirana područja na temelju vizualnih izobličenja i granu za otkrivanje sličnosti koja locira izvorna i manipulirana područja na temelju vizualnih sličnosti. Kako bi poboljšali točnost otkrivanja, Chen et al. [64] su koristili serijsku strukturu i uključili modul za pažnju s proširenom konvolucijom u fazi otkrivanja sličnosti. Struktura mreže može se vidjeti na Slici 7.1. Aria et al. [65] su predstavili metodu neovisnu o kvaliteti koja koristi generativnu suparničku mrežu za poboljšanje kvalitete slike. Također su implementirali konvolucijsku mrežu s dvije grane za otkrivanje manipulacije koja može

identificirati više manipuliranih područja i razlikovati izvor i ciljano područje manipulacije. Ova metoda je bila otporna na razne post-procesne napade i pokazala je obećavajuće rezultate u otkrivanju manipulacije na slikama niske kvalitete. Osim toga, Barni et al. [66] su predložili višedjelnu CNN mrežu koja koristi nepovratnost uzrokovanu tragovima interpolacije i nedosljednost granica manipuliranih područja kako bi precizno razlikovala između izvornih i ciljnih područja.



Slika 7.1: Serijska mreža za CMFD: (a) arhitektura predložene sheme i (b) arhitektura mreže za otkrivanje sličnosti kopiranja-premještanja. (Izvor: [64].)

7.1.2. Detekcija metode krivotvorenja spajanjem

Spajanje slika je često korištena metoda za izmjenu sadržaja slike uključujući proces kopiranja dijela s jedne slike i umetanja u drugu. Ove manipulirane slike predstavljaju značajnu prijetnju sigurnosti vizualnih informacija što zahtijeva razvoj učinkovitih metoda detekcije. Tehnike detekcije spajanja slika mogu se kategorizirati na metode ručnog odabira značajki i metode dubokog učenja. Metode temeljene na ručnom odabiru značajki obuhvaćaju tehnike temeljene na teksturi, tehnike temeljene na šumu i druge metode. Pristupi dubokog učenja automatski uče mnoštvo značajki što poboljšava točnost i generalizaciju u detekciji krivotvorenih slika. Ove metode olakšavaju izravnu optimizaciju prikaza značajki za forenziku falsifikata, potičući stalna istraživanja u ovom području. U-Net struktura posljednjih je godina dobila na važnosti za otkrivanje krivotvorina dobijenih spajanjem slika.

Wei et al. [67] predstavili su mrežu za detekciju i lokalizaciju manipulacija koja se temelji na U-Net arhitekturi s višeskalnom kontrolom (engl. *Multi-scale control*). Ova mreža koristi rotiranu rezidualnu strukturu za poboljšanje sposobnosti učenja značajki.

Slično tome, Zeng et al. [68] predložili su višenamjenski model za lociranje manipulacija spajanja u slikama koji uključuje mehanizam pažnje, gusto povezanu mrežu, ASPP i U-Net. Ovaj model učinkovito hvata višeskalne značajke (engl. *Multi-scale features*), proširuje receptivno polje i poboljšava točnost otkrivanja manipulacija spajanja u slikama. Zhang et al. [69] razvili su višenamjensku mrežu kompresije i ekstrakcije za lokalizaciju manipulacije spajanja koja se sastoji od toka vođenog maskom i toka vođenog rubom, koristeći U-Net arhitekturu. Za poboljšanje prikaza značajki, moduli za pažnju kompresije i ekscitacije (SEAM – engl. *Squeeze and excitation attention modules*) integrirani su u višenamjensku mrežu za rekalibraciju spojenih značajki. Nekoliko studija koristilo je potpuno konvolucijske mreže (FCN – engl. *Fully convolutional networks*) za detekciju krivotvorenja slika putem spajanja. Na primjer, Chen et al. [70] su uveli FCN baziranu na rezidualnim blokovima, posebno dizajniranu za lokalizaciju manipulacije spajanja slika. Ovaj pristup uključuje rezidualne blokove u FCN arhitekturu kako bi se poboljšala optimizacija mreže. Slično tome, Zhuang et al. [71] razvili su gustu FCN za detekciju manipulacija na slikama. Ovaj model koristi guste veze i proširene konvolucije za prepoznavanje suptilnih tragova manipulacije i generiranje preciznijih mapa značajki za predviđanje. Uz to, Liu et al. [72] predložili su FCN sa značajkom šuma. Ova tehnika poboljšava sposobnost generalizacije izdvajanjem značajki šuma tijekom faze predobrade, otkrivajući suptilne promjene u manipuliranim slikama. Nadalje, korištenjem mreže za predlaganje regija povećava se robusnost mreže. Ova metoda precizno locira manipulirane regije na slikama i poboljšava kako sposobnost generalizacije, tako i robusnost.

Posljednjih godina mehanizmi pažnje postali su sve napredniji i pokazali su se vrlo korisnima u području obrade prirodnog jezika. Kao rezultat toga mnogi istraživači počeli su integrirati mehanizme pažnje u detekciju manipulacija. Na primjer, Ren et al. [73] uveli su višeskalnu mrežu svjesnu konteksta s pažnjom i implementirali višeskalni višerazinski modul pažnje. Ovo ne samo da je učinkovito riješilo problem nekonzistentnih značajki na različitim skalama, već je i automatski prilagodilo težine značajki kako bi se postigla bolja reprezentacija značajki. Kako bi poboljšali točnost lokalizacije spajanja, Sun et al. [74] predložili su transformer mrežu s poboljšanim rubovima. Oni su koristili transformer mrežu s dvije grane za značajke krivotvorenja i za značajke rubova što je rezultiralo značajnim poboljšanjem u točnosti otkrivanja manipulacija.

7.1.3. Generičko otkrivanje krivotvorenja slika

Radi poboljšanja učinkovitosti otkrivanja manipulacija algoritam treba biti unaprijeđen kako bi imao više namjena. Zhang i kolege [75] su predstavili mrežu s dvije grane šuma i granica koja

koristi poboljšanu tehniku ograničene konvolucije kako bi izdvojila mape šuma, efikasno rješavajući pitanja nestabilnosti tijekom učenja. Kako bi se povećala preciznost lokalizacije uključen je modul predikcije rubova kako bi se izdvojili detalji manipuliranih rubova. Unatoč ovim poboljšanjima, učinkovitost detekcije se smanjivala kada su manipulirane slike sadržavale minimalne informacije o manipulaciji. Dong i suradnici [76] razvili su višepogledni (engl. *Multi-view*), višeskalni (engl. *Multi-scale*) nadzirani model detekcije krivotvorenja koji je objedinio granice i karakteristike šuma iz manipuliranih područja, olakšavajući učenje semantički neovisnih značajki s poboljšanom generalizacijom i točnošću detekcije. Međutim, učinkovitost detekcije se smanjivala kada su se manipulirana područja poudarala s pozadinskim područjima. Chen i suradnici [77] su predstavili mrežu koja koristi razdvajanje signala i šuma kako bi ojačala robusnost. Ovaj modul je izdvajao manipulirana područja od složenih pozadina postprocesiranjem šuma, ublažavajući negativne učinke postupaka naknadne obrade (engl. *Post-processing*) na slike i jačajući otpornost algoritma. Lin i suradnici [78] su predstavili mrežu dizajniranu za učenje s poboljšanim otkrivanjem višestrukih tragova manipulacija kombiniranjem globalnog šuma, lokalnog šuma i detaljnih karakteristika izobličenja za detekciju krivotvorenja što je rezultiralo povećanom generalizacijom i točnošću. Ipak, manipulacije s manjim izobličenjima su dovele do neotkrivenih područja manipuliranja zbog neadekvatnih tragova manipulacija. Wang i suradnici [79] su predložili višemodalni model temeljen na transformer mreži koji obuhvaća izdvajanje karakteristika visoke frekvencije, koder objekata i dekoder slika kako bi se riješio izazov prepoznavanja neprimjetnih manipulacija u RGB domeni. Inicijalno, frekvencijske karakteristike su izdvojene iz slika nakon čega su manipulirana područja identificirana spajanjem RGB i frekvencijskih karakteristika što je pokazalo učinkovitost metode na različitim skupovima podataka. Kako bi se poboljšala točnost predviđanja maski manipulacije, koristi se metoda nazvana progresivno dekodiranje maski. Liu i kolege [80] su predstavili progresivnu mrežu prostorno-kanalne korelacije (engl. *Spatio-channel correlation network*) koja se sastojala od dva puta: put odozgo prema dolje za izdvajanje lokalnih i globalnih značajki slike i put odozdo prema gore za predviđanje manipulirane maske. Modul prostorno-kanalne korelacije integriran je u mrežu kako bi dohvatio prostorne i kanalne odnose u značajkama, izvlačeći globalne tragove i poboljšavajući sposobnost mreže da se nosi s različitim napadima čime se povećava njezina otpornost. Naglašavajući problem irelevantnih semantičkih podataka, Shi i suradnici [81] su predložili postupno poboljšanu neuronsku mrežu koja je postupno identificirala manipulirana područja kroz proces grubi-do-fini (engl. *coarse-to-fine*) i koristila rotiranu rezidualnu strukturu kako bi potisnula sadržaj slike dok generira masku. Ovaj pristup je na kraju rezultirao preciznijom poboljšanom maskom.

Kako bi se suočili s izazovima vezanim uz nisku točnost otkrivanja i nepreciznu lokalizaciju granica, Gao i kolege [82] predstavili su novu dvokanalnu mrežu osjetljivu na granice, dizajniranu za otkrivanje i precizno lociranje generičkih fotografskih krivotvorenja. Ova mreža uključivala je adaptivni modul odabira frekvencija za dinamički odabir prikladnih frekvencija za izvlačenje nekonzistentnih statističkih podataka i uklanjanje nepotrebnih informacija. Također je dodan poseban modul lokalizacije granica izobličenja kako bi se poboljšala točnost lokalizacije granica. Kako bi se borili protiv problema ograničenih sposobnosti generalizacije prema neprimjetnim manipulacijama, Ganapathi i suradnici [83] predložili su model za otkrivanje fotografskih krivotvorenja temeljen na kanalnoj pažnji. Njihov pristup uključivao je modul kanalne pažnje za identificiranje i precizno lociranje manipuliranih područja koristeći međukanalne interakcije kako bi prioritizirao manipulirana područja. Slično, Xu i tim [84] su predložili međusobno komplementarnu mrežu za otkrivanje krivotvorenja koja je uključivala dva kodera za izdvajanje karakteristika prednjeg plana i pozadine. Korištenjem modula međusobne pažnje koji obuhvaća mehanizme pažnje samih karakteristika i međusobnih karakteristika, ova mreža učinkovito je uhvatila odnos između karakteristika prednjeg plana i pozadine što je rezultiralo značajnim poboljšanjem precizne lokalizacije krivotvorenih područja.

Rao (2019.) et al. [85] odlučili su poboljšati sposobnost generalizacije mrežnog modela uvodeći višesemantički uvjetni model nasumičnih polja. Ovaj model dizajniran je kako bi razlikovao krivotvorene i izvorne granice, olakšavajući identifikaciju krivotvorenih područja. Kako bi se poboljšala sposobnost mreže za prepoznavanje intrinzičnih značajki izobličenja tranzicije granica, uključeni su blokovi pažnje. Korištenjem mapa pažnje s različitim semantikama model je učinkovito spojio lokalne i globalne informacije čime je unaprijedio svoju sposobnost generalizacije. Slično tome, Li et al. [86] predložili su međudomensku mrežu pažnje sastavljenu od tri CNN mreže koje izdvajaju različite značajke poput vizualne percepcije, ponovnog uzorkovanja i lokalne nedosljednosti. Spajanje tih raznovrsnih značajki značajno je poboljšalo sposobnost modela za generalizaciju i točnu lokalizaciju krivotvorenih područja. Nadalje, Yin et al. [87] su predstavili višezadatakovnu mrežu (engl. *Multi-task network*) koja koristi učenje kontrasta za otkrivanje različitih manipulacija. Procjenom dosljednosti statističkih svojstava na različitim područjima i učenje kontrasta ojačalo je reprezentaciju značajki te posljedično poboljšalo performanse modela u zadacima otkrivanja i lokalizacije.

U rješavanju problema povezanih s niskom točnošću i nedostatkom podataka za trening, Zhou i kolege [88] osmislili su mrežu za detekciju manipulacija od grubog do finog koristeći samo-suparnički trening (engl. *Self-adversarial training*) za poboljšanje točnosti dinamičkim proširivanjem skupa podataka za učenje. Dodatno, Ren i suradnici [89] riješili su izazov ograničenih skupova podataka stvaranjem novog skupa podataka poznatog kao višestruki stvarni

skup podataka s manipuliranim scenama (engl. *multi-realistic scene manipulation dataset*). Taj skup podataka obuhvaćao je različite vrste manipulacija poput kopiranja, presijecanja i uklanjanja iz 32 različita scenarija manipulacija pronađenih u svakodnevnom životu.

Tablica 7.1 prikazuje najnovije algoritme za detekciju kopiranja i premještanja (CMFD) koji koriste tehnike dubokog učenja te prikazuje njihovu usporedbu prema vrsti detekcije. Četiri ključna aspekta obuhvaćena u Tablici 7.1 uključuju vrstu detekcije, osnovnu arhitekturu, robusnost izvedbe i skup podataka korišten u svakoj metodi.

Tablica 7.1: Usporedba pasivnih metoda otkrivanja krivotvorenja temeljenih na dubokom učenju.

Vrsta detekcije	**Godina**	**Temeljni CNN model (osnovna arhitektura)**	**Performanse robusnosti**	**Skup podataka**	**Referenca**
Kporanje-pomicanjem	2022	VGG 16, Atrous konvolucija	Promjena svjetline, zamućenje slike, JPEG kompresija, smanjenje boje, podešavanja kontrasta, dodavanje šuma	USCISI, CoMoFoD, CASIA V2	Li et al. [54]
Kporanje-pomicanjem	2022	VGG 16, SuperGlue	Rotacija, Skaliranje, Dodavanje šuma, JPEG kompresija	Self-datasets	Liu et al. [57]
Kporanje-pomicanjem	2020	DenseNet	Rotacija, Skaliranje, Dodavanje šuma, JPEG kompresija	FAU, CoMoFoD, CASIA V2	Zhong et al. [58]
Kporanje-pomicanjem	2021	VGG 16, Volterra konvolucija	Promjena svjetline, zamućenje slike, JPEG kompresija, smanjenje boje, podešavanja kontrasta, dodavanje šuma	USCISI, CoMoFoD, CASIA	Kafali et al. [59]
Kporanje-pomicanjem	2022	DenseNet, RCNN	Promjena svjetline, zamućenje slike, JPEG kompresija, smanjenje boje, podešavanja kontrasta, dodavanje šuma	CoMoFoD, MICC-F2000, CASIA V2	Nazir et al. [60]
Kporanje-pomicanjem	2022	DenseNet	Promjena svjetline, zamućenje slike, JPEG kompresija, smanjenje boje, podešavanja kontrasta, dodavanje šuma	IMD, CoMoFoD, CMHD	Zhong et al. [62]
Kporanje-pomicanjem	2018	VGG 16	Promjena svjetline, zamućenje slike, JPEG kompresija, smanjenje boje, podešavanja kontrasta, dodavanje šuma	USCISI, CoMoFoD, CASIA V2	Wu et al. [63]

Kporanje-pomicanjem	2021	VGG 16, Modul pažnje	Promjena svjetline, zamućenje slike, JPEG kompresija, smanjenje boje, podešavanja kontrasta, dodavanje šuma	USCISI, CoMoFoD, CASIA V2, COVERAGE	Chen et al. [64]
Kporanje-pomicanjem	2022	VGG 16	Promjena svjetline, zamućenje slike, JPEG kompresija, smanjenje boje, podešavanja kontrasta, dodavanje šuma	USCISI, CoMoFoD, CASIA V2	Aria et al. [65]
Kporanje-pomicanjem	2021	ResNet 50	JPEG kompresija, šum, skaliranje	SYN-Ts, USCISI, CASIA, Grip	Barni et al. [66]
Spajanje	2021	U-Net, Prstenasta rezidualna struktura	JPEG kompresija, Gaussov šum, kombinirani napad, skaliranje, rotacija	CASIA, Columbia	Wei et al. [67]
Spajanje	2022	U-Net, ASPP	JPEG kompresija, Gaussovo zamućenje	CASIA	Zeng et al. [68]
Spajanje	2021	U-Net, SEAM	JPEG kompresija, Skaliranje, Gaussovo filtriranje, Izoštravanje slike	Columbia, CASIA, Carvalho	Zhang et al. [69]
Spajanje	2020	FCN	JPEG kompresija, Gaussovo zamućenje, Gaussov šum	DVMM, CASIA, NC17, MFC18	Chen et al. [70]
Spajanje	2021	FCN	JPEG kompresija, Skaliranje	PS-scripted dataset, NIST16	Zhuang et al. [71]
Spajanje	2022	FCN, SRM	Gaussov šum, JPEG kompresija, Gaussovo zamućenje	CASIA, Columbia	Liu et al. [72]
Spajanje	2022	ResNet 50	JPEG kompresija, Gaussov šum, Skaliranje	CASIA, IMD2020, DEFACTO, SMI20K	Ren et al. [73]
Spajanje	2022	Transformer	JPEG kompresija, zamućenje medija, skaliranje	CASIA, NC2016	Sun et al. [74]
Otkivanje višestrukih manipulacija	2022	ResNet 34, Nelokalni modul	JPEG kompresija, Gaussovo zamućenje	CASIA, COVERAGE, Columbia, NIST16	Zhang et al. [75]
Otkivanje višestrukih manipulacija	2023	ResNet 50	JPEG kompresija, Gaussovo zamućenje	CASIA V2,COVERAGE, Columbia, NIST16	Dong et al. [76]
Otkivanje višestrukih manipulacija	2023	ResNet 101	JPEG kompresija, Gaussovo zamućenje, Medijan zamućenja	Self-datasets, NIST16, Columbia, CASIA	Chen et al. [77]
Otkivanje višestrukih manipulacija	2023	ResNet 50, Swin transformator	JPEG kompresija, Gaussovo zamućenje, Gaussov šum	CASIA, NIST16, Columbia, COVERAGE, CoMoFoD	Lin et al. [78]
Otkivanje višestrukih manipulacija	2022	Višemodalni transformator	JPEG kompresija, Gaussovo zamućenje,	CASIA, Columbia, Carvalho,	Wang et al. [79]

			Gaussov šum, Skaliranje	NIST16, IMD2020	
Otkivanje višestrukih manipulacija	2022	HR-Net	JPEG kompresija, skaliranje, Gaussovo zamućenje, Gaussov šum	Columbia, COVERAGE, CASIA, NIST16, IMD2020	Liu et al. [80]
Otkivanje višestrukih manipulacija	2022	VGG 19, Rotirani rezidual	JPEG kompresija, Gaussovo zamućenje, Gaussov šum	NIST16, COVERAGE, CASIA, In-The-Wild	Shi et al. [81]
Otkivanje višestrukih manipulacija	2022	ResNet 101	JPEG kompresija, Skaliranje	CASIA, Carvalho, COVERAGE, NIST16, IMD2020	Gao et al. [82]
Otkivanje višestrukih manipulacija	2022	HR-Net	Vodoravno i okomito rotiranje, zasićenje, svjetlina	CASIA V2, NIST16, Carvalho, Columbia	Ganapathi et al. [83]
Otkivanje višestrukih manipulacija	2022	VGG 16	JPEG kompresija, skaliranje, Gaussovo zamućenje, Gaussov šum	NIST16, COVERAGE, CASIA, IMD2020	Xu et al. [84]
Otkivanje višestrukih manipulacija	2022	Rezidualna jedinica, Pažnja temeljena na CRF-u, ASPP	JPEG kompresija, Skaliranje	COVERAGE, CASIA, Carvalho, IFC	Rao et al. [85]
Otkivanje višestrukih manipulacija	2022	ResNet101, Brži R-CNN	Filtriranje medijana, Gaussov šum, Gaussovo zamućenje, ponovno uzorkovanje	CASIA, Columbia, COVERAGE, NIST16	Li et al. [86]
Otkivanje višestrukih manipulacija	2022	Konvolucija i Rezidualni blok	JPEG kompresija, Gaussovo zamućenje, Gaussov šum, Skaliranje	NIST16, CASIA, COVERAGE, Columbia	Yin et al. [87]
Otkivanje višestrukih manipulacija	2022	Blok u VGG stilu	JPEG kompresija, Gaussov šum, Gaussovo zamućenje, skaliranje	DEFACTO, Columbia, CASIA, COVERAGE, NIST16	Zhou et al. [88]
Otkivanje višestrukih manipulacija	2022	ResNet 50	JPEG kompresija, Gaussov šum, Skaliranje	NIST16, CASIA, MSM30K	Ren et al. [89]

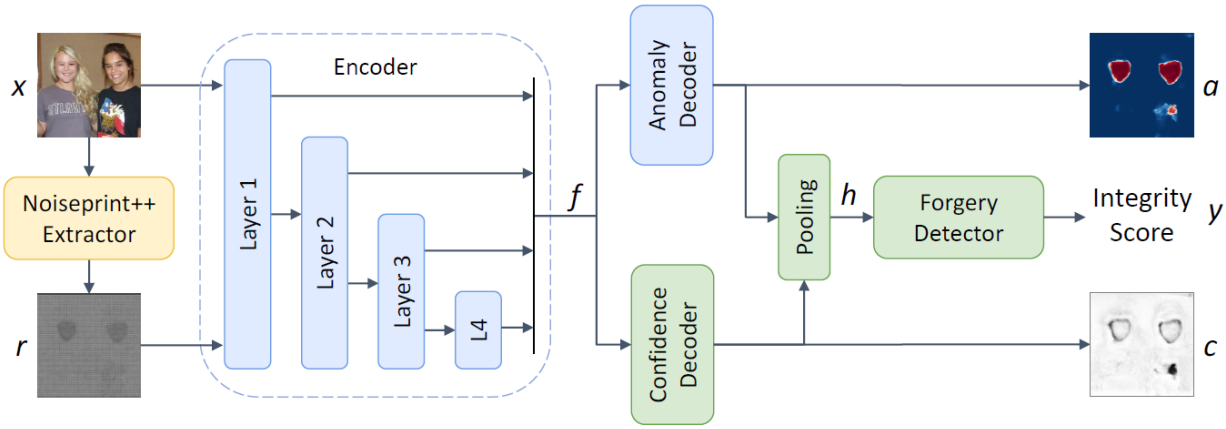
7.2. Testiranje suvremenih algoritama za otkrivanje i lokalizaciju krivotvorenja slika

Otkrivanje i lokalizacija krivotvorenja slika postaju su sve važniji u digitalnom dobu gdje manipulirane slike mogu lako širiti dezinformacije i narušiti vjerodostojnost vizualnih dokaza. U ovom poglavlju biti će testirana i predstavljena izvedba tri najsuvremenija algoritma za detekciju i lokalizaciju krivotvorenih slika: TrueFor [92], ImageForensicsOSN [93], i APSC-Net [94]. Svaki algoritam je testiran na tri široko poznata skupa podataka: CASIAv1 [95], Columbia [96] i Coverage [97], koji pružaju raznolike izazove za procjenu robusnosti i točnosti tehnika detekcije krivotvorina.

Usporebom rezultata ovih algoritama, ovo poglavlje ima za cilj istaknuti i predstaviti sveobuhvatnu analizu njihove učinkovitosti u raznim scenarijima krivotvorenja.

7.2.1. TrueFor

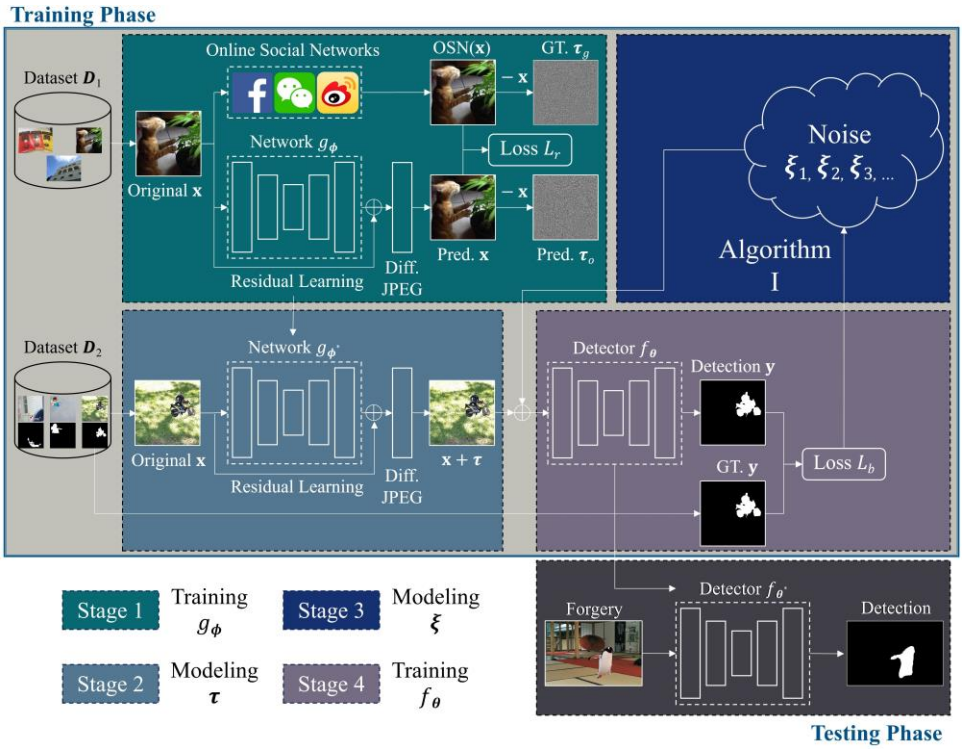
TruFor [92] je forenzički model koji se može primijeniti na širok raspon metoda manipulacije slikama, od klasičnih jednostavnih krivotvorina do novijih manipulacija temeljenih na dubokom učenju. Oslanja se na ekstrakciju visokih i niskih tragova pomoću arhitekture bazirane na transformatoru koja kombinira RGB sliku i otisak osjetljiv na šum. Algoritam uči kako ugrađivati artefakte koji se odnose na unutarnju i vanjsku obradu kamere trenirajući samo na stvarnim podacima u samokontroliranim uvjetima. Krivotvorine se otkrivaju kao odstupanja od očekivanog regularnog uzorka koji karakterizira svaku originalnu sliku. Traženje anomalija u ovom pristupu omogućava robusno otkrivanje raznih lokalnih manipulacija i osigurava generalizaciju. Uz mapu lokalizacije na razini piksela i ocjenu cjelokupnog integriteta slike, ovaj pristup daje i mapu pouzdanosti koja ističe područja u kojima predviđanja lokalizacije mogu biti sklona pogreškama. To je posebno važno u forenzičkoj primjeni kako bi se smanjila lažna upozorenja i omogućila analiza u velikom opsegu. Opsežni eksperimenti na nekoliko skupova podataka pokazuju da ova metoda pouzdano detektira i lokalizira manipulacije poput jednostavnih varki (engl. *Cheapfake*) i dubokih varki, nadmašujući neke najnovije radove u tom području.



Slika 7.2: Shema TrueFor algoritma (Izvor: [92])

7.2.2. ImageForensicsOSN

H. Wu i kolege [93] u svom su istraživanju imlali za cilj osmisliti robusnu metodu za detekciju krivotvorenja slika kako bi nadvladali obrade s gubicima (engl. *Lossy operations*) na društvenim mrežama (OSN – engl. *Online social networks*). Za rješavanje degradacija uzrokovanih OSN-ovima, predlažili su shemu modeliranja šuma i integrirali mimetičke šumove u robusni model za treniranje. OSN šumove razdvajaju na dvije komponente: 1) predvidljivi šum i 2) nepoznati šum. Prva komponenta je dizajnirana da simulira predvidljive gubitke uzrokovane poznatim obradama (npr. JPEG kompresija), čije modeliranje se oslanja na duboku neuronsku mrežu (DNN – engl. *Deep neural network*) s rezidualnim učenjem i ugrađenim diferencijalnim JPEG slojem. Druga komponenta je uglavnom odgovorna za otkrivanje nepoznatih radnji koje provode OSN-ovi i otkrivanje razlika između treniranja i testiranja različitih OSN-ova. S obzirom da je nerealno izgraditi prikladan model za nepoznati šum s aspekta samog signala, kako bi prevladali ovu poteškoću, opažanja s perspektive šuma prenose se na detektor krivotvorina, fokusirajući se samo na šum koji može uzrokovati pogoršanje performansi detekcije. Takva strategija prirodno stvara novi algoritam za modeliranje nepoznatog šuma korištenjem osnovne ideje suparničkog šuma [98], koji je u suštini neprimjetan poremećaj koji može ozbiljno narušiti performanse modela. Pokazalo se da ova robusna metoda detekcije krivotvorenja slika pokazuje vrhunsku otpornost i nadmašuje nekoliko najmodernijih algoritama.



Slika 7.3: Shema ImageForensicsOSN algoritma (Izvor: [93])

7.2.3. APSC-Net

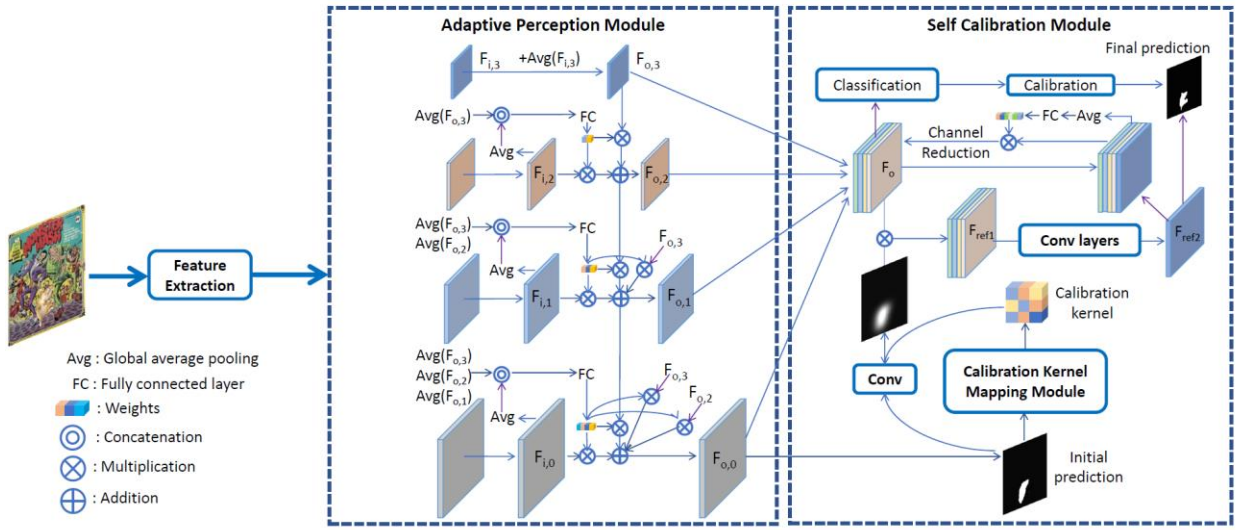
C. Qu i suradnici [94] predložili su novi model nazvan APSC-Net za preciznu lokalizaciju manipulacija na slikama. Sastoji se od ekstraktora značajki (engl. *Feature extractor*), modula za prilagodljivu percepciju (engl. *Adaptive Perception module*) i modula za samokalibraciju (eng. *Self-Calibration module*).

Modul prilagodljive percepcije - Tijekom detaljne forenzičke analize slika, ljudi često zumiraju i odzumiraju sliku, birajući optimalan skup opažanja koji im pomaže u konačnoj procjeni. Kako bi oponašali način ljudske percepcije, C. Qu i suradnici [94] dizajnirali su modul prilagodljive percepcije koji pomaže modelu usporediti različite prikaze i odabrati optimalnu kombinaciju za svaku ulaznu sliku. Ključna ideja je ponderirani zbroj trenutnih mapa značajki i svih mapa značajki više razine koristeći adaptivne težine izračunate iz njihovih globalnih prikaza.

Modul samokalibracije - Prilikom detaljne lokalizacije manipuliranih slika, ljudi su skloni potvrditi svoja početna predviđanja uspoređujući značajke koje okružuju predviđena krivotvorena područja. Također mogu izmijeniti svoja predviđanja na temelju globalne procjene autentičnosti slike. Kako bi oponašali ljudsku percepciju, C. Qu i suradnici [94] dizajnirali su modul samokalibracije za bolju izvedbu. Predloženi modul samokalibracije sastoji se od samokalibracije temeljene na segmentaciji (SSC – engl. *Segmentation-based Self Calibration*) i samokalibracije temeljene na klasifikaciji (CSC – engl. *Classification-based Self Calibration*).

Početno predviđanje za SSC se dobiva iz modula prilagodljive percepcije i unosi u mali modul za mapiranje kalibracijskog kernela, koji se sastoji od nekoliko konvolucijskih slojeva. Nakon toga se dobiva kalibracijski kernel i s njime se provodi konvolucijska operacija na početnom predviđanju. Dobivene vrijednosti se zatim normaliziraju pomoću Min-Max pristupa.

Proces za CSC započinje unošenjem poboljšanih značajki u mali klasifikacijski sloj koji predviđa je li ulazna slika manipulirana ili ne. Ako je slika predviđena kao autentična, predviđanje maske vjerojatno će imati više lažno pozitivnih (FP) rezultata pa se povećava prag binarizacije kako bi se smanjio FP. S druge strane, ako je slika predviđena kao manipulirana, smanjuje se prag binarizacije kako bi se smanjili lažno negativni rezultati. S obzirom na vjerojatnost da je ulazna slika predviđena kao krivotvorena, CSC prilagođava prag binarizacije maske za predviđanje.

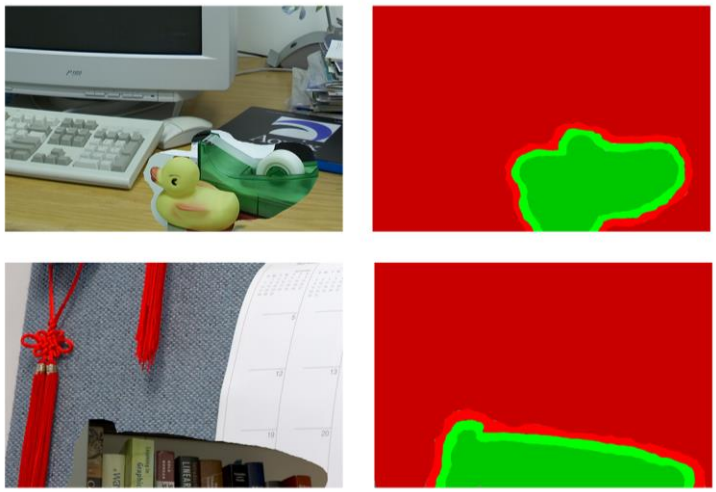


Slika 7.4: Shema APSC-Net algoritma (Izvor: [94])

7.2.4. Skupovi podataka za testiranje algoritama

Kako bi testirali algoritme za otkrivanje i lokalizaciju manipulacija na slikama korištena su tri široko poznata skupa podataka za otkrivanje krivitvorenja slika: CASIAv1, Columbia i Coverage. CASIAv1 [95] se fokusirana na tehniku krivotvorenja kopiranje-pomicanjem i spajanje te sadrži 921 krivotvorenu sliku i 800 autentičnih slika u rezoliciji 284 × 256 piksela, a slike su u JPG formatu. Columbia [96] se fokusirana na tehniku krivotvorenja spajanjem te sadrži 183 krivotvorene slike i 180 autentičnih slika u rezolicijama od 757 × 568 do 1152 × 768 piksela, a slike us u BMP i TIF formatu. Coverage [97] se fokusirana na tehniku krivotvorenja kopiranje-pomicanjem te sadrži 100 krivotvorenih slika i 100 autentičnih slika u rezoliciji 2048 × 1536 piksela, a slike su u TIF formatu.

Slika 7.5: Prikaz slike i maske iz skupa podataka za otkrivanje krivitvorenja slika iz baze CASIAv1 (Izvor: [95])



Slika 7.6: Prikaz slike i maske iz skupa podataka za otkrivanje krivitvorenja slika iz baze Columbia (Izvor: [96])

Slika 7.7: Prikaz slike i maske iz skupa podataka za otkrivanje krivitvorenja slika iz baze Coverage (Izvor: [97])

7.2.5. Rezultati testiranja algoritama za otkrivanje i lokalizaciju krivotvorenja slika

Rezultati testiranja pokazuju da APSC-Net algoritam dosljedno nadmašuje ImageForensicsOSN i TruFor na sva tri skupa podataka, a posebno se ističe na skupovima podataka CASIAv1 i Columbia. To sugerira da je APSC-Net algoritam iznimno učinkovit za lokalizaciju manipulacija na slikama, pružajući točnije i pouzdanije rezultate od uspoređenih metoda. Povišene performanse u izvedbi ističu robusnost pristupa, posebno u raznolikim i složenim scenarijima manipulacija na slikama koje predstavljaju ovi skupovi podataka.

Analiza rezultata prema skupu podataka:

1) CASIAv1 Dataset:
   - ImageForensicsOSN postiže IoU od 0.463 i F1 ocjenu od 0.508.
   - TruFor se pokazuje boljim s IoU od 0.631 i F1 ocjenom od 0.693.
   - APSC-Net značajno nadmašuje oba s IoU od 0.798 i F1 ocjenom od 0.836, pokazujući jasnu prednost u točnoj lokalizaciji manipulacija na slikama na ovom skupu podataka.
2) Coverage Dataset:
   - ImageForensicsOSN ima niže performanse s IoU od 0.182 i F1 ocjenom od 0.269.
   - TruFor pokazuje značajno poboljšanje u odnosu na ImageForensicsOSN, s IoU od 0.445 i F1 ocjenom od 0.521.
   - APSC-Net nastavlja ostvarivati najbolje rezultate s IoU od 0.488 i F1 ocjenom od 0.521, iako je razlika u poboljšanju u odnosu na TruFor manja u usporedbi s CASIAv1 skupom podataka.
3) Columbia Dataset:
   - ImageForensicsOSN postiže IoU od 0.612 i F1 ocjenu od 0.712.
   - TruFor ponovno pokazuje dobre rezultate, s IoU od 0.749 i F1 ocjenom od 0.813.
   - APSC-Net značajno nadmašuje i ImageForensicsOSN i TruFor, postižući IoU od 0.957 i F1 ocjenu od 0.968, što ukazuje na izvanrednu izvedbu na Columbia skupu podataka.

Tablica 7.2: Usporedba rezultata testiranja algoritama za otkrivanje i lokalizaciju krivotvorenja slika

Metoda	**CASIAv1 [95]**		**Coverage [97]**		**Columbia [96]**	
	IoU	F1	IoU	F1	IoU	F1
ImageForensicsOSN [93]	0.463	0.508	0.182	0.269	0.612	0.712
TruFor [92]	0.631	0.693	0.445	0.521	0.749	0.813
APSC-Net [94]	0.798	0.836	0.488	0.522	0.957	0.968

8. Zaključak

U ovom radu obrađeni su različiti aspekti tehnika otkrivanja manipulacija slikama uključujući tradicionalne pristupe temeljene na ručno izvedenim značajkama i novije napretke u metodama temeljene na dubokom učenju. Ove tehnike imaju za cilj identificirati i lokalizirati manipulirana područja unutar slika suočavajući se s izazovima poput niske točnosti, neprecizne lokalizacije granica i ograničene sposobnosti generalizacije. U posljednjih nekoliko godina istraživači su napravili značajne korake u razvoju robusnijih i preciznijih modela za otkrivanje krivotvorenja. Ti napretci uključuju integraciju inovativnih arhitektura mreža, poput mreža s dva toka i mreža pažnje kao i korištenje višezadatkovnog učenja i samo-suparnički trening za poboljšanje točnosti otkrivanja i generalizacije kroz različite scenarije manipulacije. Također su uloženi napori u stvaranje sveobuhvatnih skupova podataka koji obuhvaćaju različite vrste manipulacija pružajući istraživačima vrijedne resurse za učinkovito treniranje i evaluaciju njihovih modela. Tehnike poput postupnog dekodiranja maski i modeli višesemantičkog uvjetnog slučajnog polja uvedeni su kako bi dodatno poboljšali točnost i sposobnosti generalizacije mreža za otkrivanje krivotvorenja. Područje otkrivanja krivotvorenja na slikama nastavlja se brzo razvijati te je potaknuto napretcima u tehnikama dubokog učenja, inovativnim arhitekturama mreža i dostupnošću sveobuhvatnih skupova podataka. Ti razvoji obećavaju poboljšanje vjerodostojnosti i autentičnosti vizualnog sadržaja u različitim područjima, od digitalne forenzike do borbe protiv dezinformacija i prijevara.

9. Literatura

[1] T. Mahmood, Z. Mehmood, M. Shah, T. Saba, “A robust technique for copy-move forgery detection and localization in digital images via stationary wavelet and discrete cosine transform”, Journal of Visual Communication and Image Representation, 2018., vol. 53, str. 202–214.

[2] S.P. Jaiprakash, M.B.Desai, C.S. Prakash, V.H. Mistry, K.L. Radadiya, “Low dimensional DCT and DWT feature based model for detection of image splicing and copy-move forgery”. Multimedia Tools and Applications, 2020., vol. 79, str. 29977–30005.

[3] Y. Wo, K. Yang, G. Han, H. Chen, W. Wu, „Copy-move forgery detection based on multi-radius PCET“, IET Image Processing, 2017., vol. 11, str. 99–108.

[4] J.Y. Park, T.A. Kang, Y.H. Moon, I.K. Eom, „Copy-move forgery detection using scale invariant feature and reduced local binary pattern histogram“, Symmetry, 2020., vol. 12, str. 492.

[5] A. Rani, A. Jain, M. Kumar, „Identification of copy-move and splicing based forgeries using advanced SURF and revised template matching“, Multimedia Tools and Applications, 2021., vol. 80, str. 23877–23898.

[6] G. Singh, K. Singh, “Digital image forensic approach based on the second-order statistical analysis of CFA artifacts”, Forensic Science International: Digital Investigation, 2020, vol. 32, str. 200899.

[7] S. Taspinar, M. Mohanty, N. Memon, “PRNU-Based Camera attribution from multiple seam-Carved images”, IEEE Transactions on Information Forensics and Security, 2017, vol. 12(12):3065–80. doi: 10.1109/TIFS.2017.2737961.

[8] B. Xu, X. Wang, X. Zhou, J. Xi i S. Wang, "Source camera identification from image texture features", Neurocomputing, 2016., vol. 207, str. 131–140.

[9] Y-f. Hsu i S-f. Chang, "Camera response functions for image forensics: an automatic algorithm for splicing detection", IEEE Transactions on Information Forensics and Security, 2010., vol. 5, no. 4, str. 816–825.

[10] Y. Li i J. Zhou, "Fast and effective image copy-move forgery detection via hierarchical feature point matching", IEEE Transactions on Information Forensics and Security, 2019., vol. 14, no. 5, str. 1307–1322.

[11] S. Jaseela i S.G. Nishadha, "Copy move image forgery detection using SURF feature point extraction", International Journal of Scientific and Engineering Research, 2016., vol. 7, no. 7, str. 653–657.

[12] T. Mahmood, Z. Mehmood, M. Shah i T. Saba, "A robust technique for copy-move forgery detection and localization in digital images via stationary wavelet and discrete cosine transform", Journal of Visual Communication and Image Representation, 2018., vol. 53, str. 202–214.

[13] S. Farooq, M.H. Yousaf i F. Hussain, "A generic passive image forgery detection scheme using local binary pattern with rich models", Computers and Electrical Engineering, 2017., vol. 62, str. 459–472.

[14] A. Peng, Y. Wu i X. Kang, "Revealing traces of image resampling and resampling antiforensics", Advances in Multimedia 2017., vol. 3., str. 1–13.

[15] D. Vazquez-Padin, F. Perez-Gonzalez i P. Comesana-Alfaro, "A random matrix approach to the forensic analysis of upscaled images", IEEE Transactions on Information Forensics and Security, 2017., vol. 12, no. 9, str. 2115–2130.

[16] R.G. Van Schyndel, A.Z. Tirkel i C.F. Osborne, "A digital watermark", Proceedings of the 1st International Conference on Image Processing, Austin, TX, USA, 13–16 November 1994, IEEE, 1994., vol. 2, str. 86–90.

[17] S. Dumitrescu, X. Wu i Z. Wang, "Detection of LSB steganography via sample pair analysis", IEEE Transactions on Signal Processing, 2003., vol. 51, str. 1995–2007.

[18] H. Guo i N.D. Georganas, "Digital image watermarking for joint ownership verification without a trusted dealer", Proceedings of the International Conference on Multimedia and Expo, Baltimore, MD, USA, 6–9 July 2003, IEEE, 2003., vol. 2, str. 497–500.

[19] S.A. Parah, J.A. Sheikh, N.A. Loan i G.M. Bhat, "Robust and blind watermarking technique in DCT domain using inter-block coefficient differencing", Digital Signal Processing, 2016., vol. 53, str. 11–24.

[20] S. Etemad i M. Amirmazlaghani, "A new multiplicative watermark detector in the contourlet domain using t location-scale distribution", Pattern Recognition, 2018., vol. 77, str. 99–112.

[21] E. Etemad, S. Samavi, S. Reza Soroushmehr, N. Karimi, M. Etemad, S. Shirani i K. Najarian, "Robust image watermarking scheme using bit-plane of Hadamard coefficients", Multimedia Tools and Applications, 2018., vol. 77, str. 2033–2055.

[22] J.R. Carneiro Tavares i F. Madeiro Bernardino Junior, "Word-Hunt: a LSB steganography method with low expected number of modifications per pixel", IEEE Latin America Transactions, 2016., vol. 14, no. 2, str. 1058–1064.

[23] S.A. Laskar i K. Hemachandran, "Steganography based on random pixel selection for efficient data hiding", International Journal of Computer Engineering and Technology (IJCET), 2013., vol. 4, str. 31–44.

[24] S. Bhattacharyya, "Study and analysis of quality of service in different image based steganography using Pixel Mapping Method (PMM)", International Journal of Applied Information Systems (IJAIS), 2012., vol. 2, no. 7, str. 42–57.

[25] K. Qazanfari i R. Safabakhsh, "A new steganography method which preserves histogram: Generalization of LSB++", Information Sciences, 2014., vol. 277, str. 90–101.

[26] M. Shobana i R. Manikandan, "Efficient method for hiding data by pixel intensity", International Journal of Engineering and Technology (IJET), 2013., vol. 5, no. 1, str. 74–80.

[27] J. Ni, X. Hu i Y.Q. Shi, "Efficient JPEG steganography using domain transformation of embedding entropy", IEEE Signal Processing Letters, 2018., vol. 25, no. 6, str. 773–777.

[28] N. Ghoshal i J.K. Mandal, "A steganographic scheme for colour image authentication (SSCIA)", Proceedings of the International Conference on Recent Trends in Information Technology (ICRTIT), 2011, str. 826–831.

[29] A. Ibaida i I. Khalil, "Wavelet-Based ECG steganography for protecting patient confidential information in point-of-care systems", IEEE Transactions on Biomedical Engineering, 2013., vol. 60, no. 12, str. 3322–3330.

[30] H. Al-dmour i A. Al-ani, "A steganography embedding method based on edge identification and XOR coding", Expert Systems with Applications, 2016., vol. 46, str. 293–306.

[31] S.E. Jero i P. Ramu, "Curvelets-based ECG steganography for data security", Electronics Letters, 2016., vol. 52, no. 4, str. 283–285.
[32] M. Urvoy, D. Goudia i F. Autrusseau, "Perceptual DFT watermarking with improved detection and robustness to geometrical distortions", IEEE Transactions on Information Forensics and Security, 2014., vol. 9, no. 7, str. 1108–1119.
[33] A. Bamatraf, R. Ibrahim, M. Najib i M. Salleh, "A new digital watermarking algorithm using combination of least significant bit (LSB) and inverse bit", Journal of Computer Science, 2011., vol. 3, no. 4, str. 2151–2167.
[34] A. Bose i S.P. Maity, "Spread spectrum watermark detection on degraded compressed sensing", IEEE Sensors Letters, 2017., vol. 1, no. 5, str. 1–4.
[35] K.S. Bapat i Radhika V. Totla, "Comparative analysis of watermarking in digital images using DCT & DWT", International Journal of Scientific Research and Publications (IJSRP), 2013., vol. 3, no. 2., str. 1–4.
[36] F. Ernawan i M.N. Kabir, "A robust image watermarking technique with an optimal DCT-Psychovisual threshold", IEEE Access, 2018., vol. 6, str. 20464–20480.
[37] B.E. Khoo, N.M. Makbol i T.H. Rassem, "Block-based discrete wavelet transform-singular value decomposition image watermarking scheme using human visual system characteristics", IET Image Processing, 2016., vol. 10, no. 1, str. 34–52.
[38] Y. Rao i J. Ni, "A deep learning approach to detection of splicing and copy-move forgeries in images", Proceedings of the 2016 IEEE International Workshop on Information Forensics and Security (WIFS), 2016, str. 1–6.
[39] N.H. Rajini, "Image forgery identification using convolution neural network", International Journal of Recent Technology and Engineering, 2019., vol. 8, str. 311–320.
[40] J. Fridrich, D. Soukal i J. Lukás, "Detection of copy move forgery in digital images", Proceedings of the Digital Forensic Research Workshop, 2003.
[41] A.C. Popescu i H. Farid, "Exposing digital forgeries by detecting duplicated image regions", Technical Report TR2004-515, 2004.
[42] H. Farid, "Detecting digital forgeries using bispectral analysis", Technical Report AIM-1657, AI Laboratory, Massachusetts Institute of Technology, 1999.
[43] D. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, 2004., vol. 60, str. 91–110.
[44] H. Bay, A. Ess, T. Tuytelaars i L. Van Gool, "Speeded-up robust features (SURF)", Computer Vision and Image Understanding, 2008., vol. 110, no. 3, str. 346–359.
[45] E. Rublee, V. Rabaud, K. Konolige i G. Bradski, "ORB: an efficient alternative to SIFT or SURF", Proceedings of the 2011 International Conference on Computer Vision (ICCV), 2011, str. 2564–2571.
[46] J. Lukás i J. Fridrich, "Estimation of primary quantization matrix in double compressed JPEG images", Proceedings of the Digital Forensic Research Workshop, 2003.
[47] J. Fridrich, M. Chen i M. Goljan, "Imaging sensor noise as digital x-ray for revealing forgeries", Proceedings of the 9th International Workshop on Information Hiding, Saint-Malo, France, 2007, str. 342–358.
[48] M.K. Johnson i H. Farid, "Exposing digital forgeries through chromatic aberration", Proceedings of the ACM Multimedia and Security Workshop, Geneva, 2006, str. 48–55.
[49] M. Fischler i R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography", Communications of the ACM, 1981., vol. 24, no. 6, str. 381–395.

[50] M.K. Johnson i H. Farid, "Detecting photographic composites of people", Proceedings of the 6th International Workshop on Digital Watermarking, Guangzhou, 2007, str. 19-33.

[51] M.K. Johnson i H. Farid, "Metric measurements on a plane from a single image", Technical Report TR2006-579, 2006.

[52] Y. Rao i J. Ni, "A deep learning approach to detection of splicing and copy-move forgeries in images", Proceedings of the IEEE International Workshop on Information Forensics and Security, Abu Dhabi, United Arab Emirates, 4–7 December 2016, str. 1–6.

[53] S. Kumar i S.K. Gupta, "A robust copy move forgery classification using end-to-end convolution neural network", Proceedings of the 8th International Conference on Reliability, Infocom Technologies and Optimization, Noida, India, 4–5 June 2020, str. 253–258.

[54] Q. Li, C. Wang, X. Zhou i Z. Qin, "Image copy-move forgery detection and localization based on super-BPD segmentation and DCNN", Scientific Reports 2022, vol. 12, br. članka 14987.

[55] J. Wan, Y. Liu, D. Wei, X. Bai i Y. Xu, "Super-BPD: Super boundary-to-pixel direction for fast image segmentation", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020, str. 9250–9259.

[56] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy i A.L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018., vol. 40, str. 834–848.

[57] Y. Liu, C. Xia, X. Zhu i S. Xu, "Two-stage copy-move forgery detection with self deep matching and proposal superglue", IEEE Transactions on Image Processing, 2021., vol. 31, str. 541–555.

[58] J.L. Zhong i C.M. Pun, "An end-to-end Dense-InceptionNet for image copy-move forgery detection", IEEE Transactions on Information Forensics and Security, 2019., vol. 15, str. 2134–2146.

[59] E. Kafali, N. Vretos, T. Semertzidis i P. Daras, "RobusterNet: Improving copy-move forgery detection with Volterra-based convolutions", Proceedings of the 25th International Conference on Pattern Recognition, Milan, Italy, 10–15 January 2021, IEEE: New York, NY, USA, 2021; str. 1160–1165.

[60] T. Nazir, M. Nawaz, M. Masood i A. Javed, "Copy move forgery detection and segmentation using improved mask region-based convolution network (RCNN)", Applied Soft Computing, 2022., vol. 131, str. 109778.

[61] K. He, G. Gkioxari, P. Dollár i R. Girshick, "Mask R-CNN", Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017, str. 2980–2988.

[62] J.L. Zhong, J.X. Yang, Y.F. Gan, L. Huang i H. Zeng, "Coarse-to-fine spatial-channel-boundary attention network for image copy-move forgery detection", Soft Computing, 2022., vol. 26, str. 11461–11478.

[63] Y. Wu, W. Abd-Almageed i P. Natarajan, "Busternet: Detecting copy-move image forgery with source/target localization", Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018, str. 168–184.

[64] B. Chen, W. Tan, G. Coatrieux, Y. Zheng i Y.Q. Shi, "A serial image copy-move forgery localization scheme with source/target distinguishment", IEEE Transactions on Multimedia, 2020., vol. 23, str. 3506–3517.

[65] M. Aria, M. Hashemzadeh i N. Farajzadeh, "QDL-CMFD: A quality-independent and deep learning-based copy-move image forgery detection method", Neurocomputing, 2022., vol. 511, str. 213–236.

[66] M. Barni, Q.T. Phan i B. Tondi, "Copy move source-target disambiguation through multi-branch CNNs", IEEE Transactions on Information Forensics and Security, 2020., vol. 16, str. 1825–1840.

[67] Y. Wei, Z. Wang, B. Xiao, X. Liu, Z. Yan i J. Ma, "Controlling neural learning network with multiple scales for image splicing forgery detection", ACM Transactions on Multimedia Computing, Communications, and Applications, 2020., vol. 16, str. 1–22.

[68] P. Zeng, L. Tong, Y. Liang, N. Zhou i J. Wu, "Multitask image splicing tampering detection based on attention mechanism", Mathematics, 2022., vol. 10, str. 3852.

[69] Y. Zhang, G. Zhu, L. Wu, S. Kwong, H. Zhang i Y. Zhou, "Multi-task SE-network for image splicing localization", IEEE Transactions on Circuits and Systems for Video Technology, 2022., vol. 32, str. 4828–4840.

[70] B. Chen, X. Qi, Y. Zhou, G. Yang, Y. Zheng i B. Xiao, "Image splicing localization using residual image and residual-based fully convolutional network", Journal of Visual Communication and Image Representation, 2020., vol. 73, str. 102967.

[71] P. Zhuang, H. Li, S. Tan, B. Li i J. Huang, "Image tampering localization using a dense fully convolutional network", IEEE Transactions on Information Forensics and Security, 2021., vol. 16, str. 2986–2999.

[72] Q. Liu, H. Li i Z. Liu, "Image forgery localization based on fully convolutional network with noise feature", Multimedia Tools and Applications, 2022., vol. 81, str. 17919–17935.

[73] R. Ren, S. Niu, J. Jin, J. Zhang, H. Ren i X. Zhao, "Multi-scale attention context-aware network for detection and localization of image splicing", Applied Intelligence, 2023., vol. 53, str. 18219–18238.

[74] Y. Sun, R. Ni i Y. Zhao, "ET: Edge-enhanced transformer for image splicing detection", IEEE Signal Processing Letters, 2022., vol. 29, str. 1232–1236.

[75] Z. Zhang, Y. Qian, Y. Zhao, X. Zhang , L. Zhu, J. Wang i J. Zhao, "Noise and edge based dual branch image manipulation detection", CNIOT '23: Proceedings of the 2023 4th International Conference on Computing, Networks and Internet of Things, 2023., str. 963-968.

[76] C. Dong, X. Chen, R. Hu, J. Cao i X. Li, "MVSS-Net: Multi-view multi-scale supervised networks for image manipulation detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023., vol. 45, str. 3539–3553.

[77] J. Chen, X. Liao, W. Wang, Z. Qian, Z. Qin i Y. Wang, "SNIS: A signal noise separation-based network for post-processed image forgery detection", IEEE Transactions on Circuits and Systems for Video Technology, 2023., vol. 33, str. 935–951.

[78] X. Lin, S. Wang, J. Deng, Y. Fu, X. Bai, X. Chen, X. Qu i W. Tang, "Image manipulation detection by multiple tampering traces and edge artifact enhancement", Pattern Recognition, 2023., vol. 133, str. 109026.

[79] J. Wang, Z. Wu, J. Chen, X. Han, A. Shrivastava, S.N. Lim i Y.G. Jiang, "Objectformer for image manipulation detection and localization", Proceedings of the Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; IEEE: New York, NY, USA, 2022., str. 2354–2363.

[80] X. Liu, Y. Liu, J. Chen i X. Liu, "PSCC-Net: Progressive spatio-channel correlation network for image manipulation detection and localization", IEEE Transactions on Circuits and Systems for Video Technology, 2022., vol. 32, str. 7505–7517.

[81] Z. Shi, C. Chang, H. Chen, X. Du i H. Zhang, "PR-Net: Progressively-refined neural network for image manipulation localization", International Journal of Intelligent Systems, 2022., vol. 37, pp. 3166–3188.

[82] Z. Gao, C. Sun, Z. Cheng, W. Guan, A. Liu i M. Wang, "TBNet: A two-stream boundary-aware network for generic image manipulation localization", IEEE Transactions on Knowledge and Data Engineering, 2023., vol. 35, str. 7541–7556.

[83] I.I. Ganapathi, S. Javed, S.S. Ali, A. Mahmood, N.S. Vu i N. Werghi, "Learning to localize image forgery using end-to-end attention network", Neurocomputing, 2022., vol. 512, str. 25–39.

[84] D. Xu, X. Shen, Y. Lyu, X. Du i F. Feng, "MC-Net: Learning mutually-complementary features for image manipulation localization", International Journal of Intelligent Systems, 2022., vol. 37, str. 3072–3089.

[85] Y. Rao, J. Ni i H. Xie, "Multi-semantic CRF-based attention model for image forgery detection and localization", Signal Processing, 2021., vol. 183, str. 108051.

[86] S. Li, S. Xu, W. Ma i Q. Zong, "Image manipulation localization using attentional cross-domain CNN features", IEEE Transactions on Neural Networks and Learning Systems, 2021., str. 1–15.

[87] Q. Yin, J. Wang, W. Lu i X. Luo, "Contrastive learning based multi-task network for image manipulation detection", Signal Processing, 2022., vol. 201, str. 108709.

[88] L. Zhuo, S. Tan, B. Li i J. Huang, "Self-adversarial training incorporating forgery attention for image forgery localization", IEEE Transactions on Information Forensics and Security, 2022., vol. 17, str. 819–834.

[89] R. Ren, S.Niu, H. Ren, S. Zhang, T. Han i X. Tong, „ESRNet: Efficient search and recognition network for image manipulation detection“, ACM Transactions on Multimedia Computing, Communications, and Applications, 2022., vol. 18, str. 1–23.

[90] W.D. Ferreira , C.B.R. Ferreira, G. da Cruz Júnior i F. Soares, „A review of digital image forensics“, Computers and Electrical Engineering, 2020., vol. 85, str. 106685.

[91] C. Shi, L. Chen, C. Wang, X. Zhou i Z. Qin, „Review of Image Forensic Techniques Based on Deep Learning“, Mathematics, 2023., vol. 11, str. 3134.

[92] F. Guillaro, D. Cozzolino, A. Sud, N. Dufour i L. Verdoliva, „TruFor: Leveraging All-Round Clues for Trustworthy Image Forgery Detection and Localization“, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023., str. 20606-20615.

[93] H. Wu, J. Zhou, J. Tian i J. Liu, „Robust Image Forgery Detection over Online Social Network Shared Images“, Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), 2022., str. 13430-13439.

[94] C. Qu, Y. Zhong, C. Liu, G. Xu, D. Peng, F. Guo i L. Jin, „Towards Modern Image Manipulation Localization: A Large-Scale Dataset and Novel Methods“, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024., str. 10781-10790.

[95] J. Dong, W. Wang i T. Tan, „Casia image tampering detection evaluation database“, Proceedings of the IEEE China Summit and International Conference on Signal and Information Processing, Beijing, China, 6–10 July, 2013.; IEEE: New York, NY, USA, 2013, str. 422–426.

[96] Y.F. Hsu i S.F. Chang, „Detecting image splicing using geometry invariants and camera characteristics consistency“, Proceedings of the IEEE International Conference on Multimedia and Expo, Toronto, ON, Canada, 9–12 July, 2006.; IEEE: New York, NY, USA, 2006., str. 549–552.

[97] B. Wen, Y. Zhu, R. Subramanian, T.T. Ng, X. Shen i S. Winkler, COVERAGE–A novel database for copy-move forgery detection. In Proceedings of the IEEE International Conference on Image Processing, Phoenix, AZ, USA, 25–28 September, 2016.; IEEE: New York, NY, USA, 2016., str. 161–165.

[98] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow i B. Fergus, “Intriguing properties of neural networks”, Proceedings International Conference on Learning Representations, 2014., str. 1-10.

[99] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo i G. Serra, „A SIFT-based forensic method for copy-move attack detection and transformation recovery“, IEEE Transactions on Information Forensics and Security 2011., vol. 6, str. 1099–1110.

[100] T.J. de Carvalho, C. Riess, E. Angelopoulou, H.Pedrini i A. de Rezende Rocha, „Exposing digital image forgeries by illumination color classification“. IEEE Transactions on Information Forensics and Security, 2013, vol. 8, str. 1182–1194.

[101] D. Tralic, I. Zupancic, S. Grgic i M. Grgic, „CoMoFoD–New database for copy-move forgery detection“, Proceedings of the International Symposium Electronics in Marine, Zadar, Croatia, 13–15 September, 2013.; IEEE: New York, NY, USA, 2013., str. 49–54.

[102] P. Korus, „Digital image integrity–A survey of protection and verification techniques“, Digital Signal Processing, 2017., vol. 71, str. 1–26.

[103] Y. Wu, W. Abd-Almageed i P. Natarajan, „Busternet: Detecting copy-move image forgery with source/target localization“, Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September, 2018., str. 168–184.

[104] H. Guan, M. Kozak, E. Robertson, Y. Lee, A.N. Yates, A. Delgado, D. Zhou, T. Kheyrkhah, J. Smith i J. Fiscus, „MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation“, Proceedings of the IEEE Winter Applications of Computer Vision Workshops,Waikoloa, HI, USA, 7–11 January, 2019.; IEEE: New York, NY, USA, 2019., str. 63–72.

[105] G. Mahfoudi, B. Tajini, F. Retraint, F. Morain-Nicolier, J.L. Dugelay i P. Marc, „DEFACTO: Image and face manipulation dataset“, Proceedings of the 27th European Signal Processing Conference, A Coruna, Spain, 2–6 September, 2019.; IEEE: New York, NY, USA, 2019., str. 1–5.

[106] A. Novozamsky, B. Mahdian i S. Saic, „IMD2020: A large-scale annotated dataset tailored for detecting manipulated images“, Proceedings of the IEEE/CVFWinter Conference on Applications of Computer VisionWorkshops, Snowmass, CO, USA, 1–5 March, 2020.; IEEE: New York, NY, USA, 2020., str. 71–80.

10. Popis slika

Sveučilište
Sjever

SVEUČILIŠTE
SJEVER

IZJAVA O AUTORSTVU

Završni/diplomski/specijalistički rad isključivo je autorsko djelo studenta koji je isti izradio te student odgovara za istinitost, izvornost i ispravnost teksta rada. U radu se ne smiju koristiti dijelovi tuđih radova (knjiga, članaka, doktorskih disertacija, magistarskih radova, izvora s interneta, i drugih izvora) bez navođenja izvora i autora navedenih radova. Svi dijelovi tuđih radova moraju biti pravilno navedeni i citirani. Dijelovi tuđih radova koji nisu pravilno citirani, smatraju se plagijatom, odnosno nezakonitim prisvajanjem tuđeg znanstvenog ili stručnoga rada. Sukladno navedenom studenti su dužni potpisati izjavu o autorstvu rada.

Ja, SPOMENKO KEŠINA *(ime i prezime)* pod punom moralnom, materijalnom i kaznenom odgovornošću, izjavljujem da sam isključivi autor/ica završnog/diplomskog/specijalističkog (*obrisati nepotrebno*) rada pod naslovom OTKRIVANJE KRIVOTVORENJA SLIKA I NJIHOVA LOKALIZACIJA *(upisati naslov)* te da u navedenom radu nisu na nedozvoljeni način (bez pravilnog citiranja) korišteni dijelovi tuđih radova.

Student/ica:
(upisati ime i prezime)

S. Kešina
(vlastoručni potpis)

Sukladno članku 58., 59. i 61. Zakona o visokom obrazovanju i znanstvenoj djelatnosti završne/diplomske/specijalističke radove sveučilišta su dužna objaviti u roku od 30 dana od dana obrane na nacionalnom repozitoriju odnosno repozitoriju visokog učilišta.

Sukladno članku 111. Zakona o autorskom pravu i srodnim pravima student se ne može protiviti da se njegov završni rad stvoren na bilo kojem studiju na visokom učilištu učini dostupnim javnosti na odgovarajućoj javnoj mrežnoj bazi sveučilišne knjižnice, knjižnice sastavnice sveučilišta, knjižnice veleučilišta ili visoke škole i/ili na javnoj mrežnoj bazi završnih radova Nacionalne i sveučilišne knjižnice, sukladno zakonu kojim se uređuje umjetnička djelatnost i visoko obrazovanje.