دوفصل‌نامه علمی منادی امنیت فضای تولید و تبادل اطلاعات (افتا)
سال ۱۴۰۲، جلد ۲۱، شماره ۲ (۴۲-۵۰)
https://monadi.isc.org.ir

ارائه‌شده در بیستیمین کنفرانس بین‌المللی انجمن رمز ایران در سازمان پژوهش‌های علمی و صنعتی ایران، ۸ و ۹ شهریور ۱۴۰۲

بررسی جامع کارایی روش نهان‌کاوی مبتنی بر یادگیری عمیق در کشف روش‌های حوزه مکان★

وجیهه ثابتی* و مهدیه سمیعی ولوجردی

گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه الزهرا، تهران، ایران

اطلاعات مقاله

کلمات کلیدی:
نهان‌نگاری
نهان‌کاوی
نهان‌نگاری مبتنی بر حوزه مکان
یادگیری عمیق
شبکه عصبی پیچشی

dor: 20.1001.1.24763047.1402.12.2.5.1

نوع مقاله: پژوهشی

چکیده

نهان‌نگاری، هنر مکاتبات پنهانی است که در آن یک پیام به صورت مخفیانه منتقل می‌شود و نهان‌کاوی، هنر کشف حضور اطلاعات پنهان است. شبکه‌های عصبی پیچشی برخلاف روش‌های نهان‌کاوی سنتی، با استخراج خودکار ویژگی‌ها، وجود داده را تشخیص می‌دهند. در مقالات مختلف، عملکرد مدل‌های موجود بر روی تعداد محدودی از روش‌های نهان‌نگاری حوزه مکان گزارش شده است. هدف اصلی این مقاله، ارائه یک شبکه عصبی پیچشی و بررسی جامع عملکرد آن در کشف روش‌های حوزه مکان مختلف است. مدل پیشنهادی از سه قسمت پیش‌پردازش، ماژول پیچشی و طبقه‌بند تشکیل شده است. در لایه ماژول پیچشی از ۵ بلوک و طبقه‌بند از ۳ لایه تمام متصل تشکیل شده است. از روش‌های جاسازی در بیت کم ارزش، جاسازی در مقدار اختلاف پیکسل‌ها و جاسازی مبتنی بر ایده تطبیقی برای تست استفاده شده است. روش پیشنهادی می‌تواند با دقت بالاتر از ۹۷٪ وجود داده‌های با طول‌های حتی بسیار کم در روش‌های دو گروه اول را شناسایی کند. عملکرد روش پیشنهادی در کشف درصد جاسازی‌های بسیار کم روش تطبیقی با دقت بالای ۷۰٪ بسیار مناسب است و این ویژگی نقطه تمایز مدل پیشنهادی نسبت به روش‌های سنتی است. زیرا موفقیت روش‌های استخراج ویژگی دستی به دلیل کم بودن تغییرات ویژگی‌های آماری در سطوح جاسازی پایین، بسیار کمتر است.

۱ مقدمه

با توسعه شبکه‌های اجتماعی و استفاده از اینترنت به عنوان کانال انتقال اطلاعات، امنیت اطلاعات به یک چالش بسیار مهم تبدیل شده است. تا به حال روش‌های متعددی به منظور دست‌یابی به انتقال ایمن اطلاعات، ارائه شده است. رمزنگاری برای محافظت از محتوای پیام، ته‌نقش‌نگاری برای حفظ حق کپی و نهان‌نگاری [1] برای پنهان کردن وجود ارتباط استفاده می‌شود. در بسیاری از کاربردها، محرمانه بودن ارتباطات برای جلوگیری از دسترسی غیرقانونی بسیار ضروری است، مانند حوزه پزشکی، نظامی، ارتباطات سازمان‌های اطلاعاتی، سیستم اطلاعات شخصی و ...؛ بنابراین نهان‌نگاری کاربرد بالقوه زیادی در ارتباطات ایمن محیط داده‌های بسیار بزرگ دارد [۱].

هر روش نهان‌نگاری مبتنی بر تصویر، با استفاده از یک الگوریتم جاسازی، داده سرّی را در تصویر پوشانه [2] پنهان می‌کند. خروجی هر الگوریتم جاسازی، یک تصویر گنجانه [3] است. جاسازی داده در دو

[1] Steganography [2] Cover image [3] Stego image

★از کمیته علمی بیستمین کنفرانس بین‌المللی انجمن رمز ایران برای داوری این مقاله تشکر می‌شود.
*نویسنده مسئول.
آدرس‌های رایانامه: v.sabeti@alzahra.ac.ir (وجیهه ثابتی)، samiee.mahdis@yahoo.com (مهدیه سمیعی ولوجردی)

حوزه مکان یا تبدیل انجام می‌شود. در روش‌های حوزه مکان، با تغییر مستقیم تعدادی از پیکسل‌های تصویر پوشانه، جاسازی داده انجام می‌شود. جاسازی در بیت‌های کم ارزش پیکسل‌ها (LSB [۱]) (مانند روش‌های LSBF [۲] و LSBM [۳]) و جاسازی در مقدار تفاوت زوج پیکسل‌ها (PVD [۴]) [۲] از ساده‌ترین ایده‌های روش‌های مکانی است. ایده دیگر در این حوزه، جاسازی تطبیقی، یعنی جاسازی با توجه به نواحی اطراف پیکسل است که به دلیل مطابقت با ویژگی‌های سیستم بینایی انسان، این روش‌ها در مقابل روش‌های نهان‌کاوی امنیت بیشتری دارند. روش تطبیقی بر مبنای پیچیدگی (CBL [۵]) یکی از روش‌های تطبیقی موفق است [۳]. از دیگر روش‌های این حوزه می‌توان به HUGO [۴]، HILL [۵]، MiPOD [۶]، S-UNIWARD [۷] و WOW [۸] اشاره کرد.

نهان‌کاوی، مهارت کشف وجود داده‌های پنهان است. در اغلب روش‌های جامع [۶]، نهان‌کاوی به عنوان یک مسئله رده‌بندی دودویی فرموله می‌شود. هدف این روش‌ها کشف تصاویر گنجانه است و برخلاف روش‌های خاص، هیچ وابستگی به روش جاسازی داده ندارند [۹]. در روش‌های جامع، ابتدا ویژگی‌های دستی از مجموعه تصاویر آموزش استخراج می‌شود و سپس یک رده‌بند دودویی مناسب مانند SVM [۱۰] یا رده‌بند جمعی (EC [۷]) [۱۱] براساس ویژگی‌های استخراج شده آموزش می‌یابد. از این رده‌بند آموزش دیده می‌توان برای تصمیم‌گیری در مورد پوشانه یا گنجانه بودن یک تصویر جدید استفاده کرد. SPAM [۸] [۱۲]، SRM [۹] [۱۳] و maxSRM [۱۰] [۱۴] از بردارهای ویژگی موفق موجود هستند.

میزان کارایی کم، زیاد بودن ابعاد و نیاز محاسباتی بالا از مهم‌ترین چالش‌های این روش‌های نهان‌کاوی است. هدف اصلی استفاده روش‌های نهان‌کاوی از شبکه‌های عصبی پیچشی (CNN [۱۱])، غلبه بر بخشی از این چالش‌ها است. اگر چه آموزش این شبکه‌ها به داده زیاد و سخت‌افزار نیاز دارد، اما پیشرفت‌های اخیر در زمینه شبکه عصبی، قدرت محاسباتی ارائه شده توسط GPU [۱۲] و فراوانی داده‌ها، از دلایل فراگیر شدن موفقیت‌آمیز استفاده از روش‌های یادگیری عمیق در هوش مصنوعی و حل مسائل پردازش تصویر است [۱۵].

CNNهایی که اخیرا مورد مطالعه و تحقیق در زمینه نهان‌کاوی قرار گرفته‌اند، می‌توانند ویژگی‌های موجود در تصاویر را به صورت خودکار استخراج کرده و تصاویر پوشانه و گنجانه را طبقه‌بندی کنند [۱۶]. برخی از این روش‌ها با هدف کشف روش‌های نهان‌نگاری حوزه مکان ارائه شده‌اند. اما بررسی مقالات موجود نشان می‌دهد که اغلب آنها در مرحله تست، دقت خود در کشف روش‌های بسیار معدودی را گزارش کرده‌اند، در حالی که روش‌های مکانی بسیار زیاد با ایده‌های متفاوتی وجود دارد که تا به حال عملکرد روش‌های مبتنی بر CNN در برابر آنها گزارش نشده است. در این مقاله، روش Yedroudj-Net [۱۷] به عنوان یکی از روش‌های مبتنی بر CNN موفق در حوزه مکان انتخاب شده است. پس از بررسی، مشخص شد که در این مدل بیش‌برازش [۱۳] زودهنگام رخ می‌دهد. در این مقاله اصلاحی برای این مدل پیشنهاد می‌شود. برای سنجش عملکرد این معماری جدید، از روش‌های مکانی موجود در سه گروه مختلف استفاده می‌شوند. به علاوه، عملکرد روش پیشنهادی با تعدادی از روش‌های نهان‌کاوی سنتی موجود نیز مقایسه می‌شود. به طور خلاصه، مهم‌ترین اهداف این مقاله عبارتند از:

- اصلاح مدل Yedroudj-Net در جهت به تاخیر انداختن پدیده بیش‌برازش
- بررسی عملکرد مدل پیشنهادی در برابر سه گروه از روش‌های مکانی
- مقایسه عملکرد مدل پیشنهادی با روش‌های مرسوم استخراج دستی ویژگی‌ها

در ادامه، در بخش ۲، تعدادی از مدل‌های مبتنی بر یادگیری عمیق پیشنهاد شده برای کشف روش‌های نهان‌نگاری مکانی معرفی می‌شوند. در بخش ۳، مدل روش پیشنهادی با توجه به مدل پایه شرح داده می‌شود. در بخش ۴، نتایج به دست آمده از جمله مقدار دقت و خطای مدل پیشنهادی در کشف تعدادی از روش‌های مکانی در مقایسه با مدل‌های سنتی ارائه می‌شود. در انتها، نتیجه‌گیری و پیشنهاداتی برای ادامه کار بیان می‌شود.

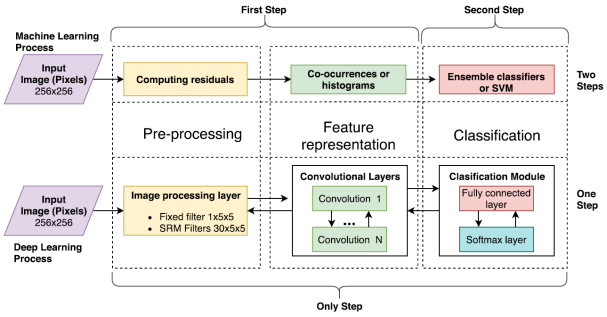
۲ مرور کارهای مرتبط

نهان‌کاوری، علم حمله به نهان‌نگاری در نبردی است که هرگز پایان نمی‌یابد و هدف اصلی آن جمع‌آوری مدارک و شواهد کافی برای اثبات حضور پیام مخفی جاسازی شده و شکستن امنیت حامل آن پیام است. روش‌های نهان‌کاوری مختلفی تا به حال پیشنهاد شده است و در مقالات مختلف، دسته‌بندی‌های متفاوتی برای این روش‌ها ارائه شده است. در یکی از این دسته‌بندی‌ها، روش‌های پنهان‌شکنی در سه دسته بصری، ساختاری یا آماری طبقه‌بندی می‌شوند که ایده اصلی آنها به ترتیب شناسایی ناهنجاری‌های بصری داخل تصویر گنجانه، شناسایی تغییرات ایجاد شده در فرمت فایل و تجزیه و تحلیل ویژگی‌های آماری شیء گنجانه در مقایسه با مجموعه‌ای از پوشانه‌ها است. از دیدگاه دیگر، روش‌های نهان‌کاوی می‌توانند خاص یا جامع باشند. روش‌های نهان‌کاوی خاص برای یک الگوریتم نهان‌نگاری خاص طراحی شده‌اند، در حالی که نهان‌کاوی جامع یا کور، یک تکنیک کلی است، که می‌تواند داده جاسازی شده توسط هر الگوریتم نهان‌نگاری، حتی یک الگوریتم ناشناخته را کشف کند.

بیشتر تحقیقات اخیر در حوزه نهان‌کاوی بر روی روش‌های جامع متمرکز شده‌اند. ویژگی مثبت این روش‌ها، قابلیت استفاده آنها برای روش‌های جاسازی مختلف می‌باشد. در حال حاضر یکی از رویکردها برای

[1] Least Significant Bit (LSB) [2] LSB Flipping (LSBF) [3] LSB Matching (LSBM) [4] Pixel Value Differencing (PVD) [5] Complexity Based LSBM (CBL) [6] Universal methods [7] Ensembler Classifier (EC) [8] Subtractive Pixel Adjacent Matrix (SPAM) [9] Spatial Rich Model (SRM) [10] Max Spatial Rich Model (maxSRM) [11] Convolutional Neural Network (CNN) [12] Graphics Processing Unit (GPU)

[13] Overfitting



شکل ۱. ساختار کلی روش‌های نهان‌کاوی سنتی با استخراج دستی ویژگی‌ها (بخش بالا) و روش‌های مبتنی بر یادگیری عمیق (بخش پایین)

بهبود کارایی الگوریتم‌های نهان‌کاوی کور، ارتقای کیفیت بردار ویژگی‌های استخراج شده از تصاویر است. هرچه این بردار غنی‌تر باشد و ویژگی‌های مناسب‌تری را در بر بگیرد، عملکرد الگوریتم بهتر خواهد بود. بنابراین کمیت و کیفیت ویژگی‌های استخراج شده از تصویر، به یک فاکتور مهم در طراحی نهان‌کاوی کور تبدیل شده است. بهترین تکنیک‌های نهان‌کاوی سنتی همگی از یک خط مشترک استفاده می‌کنند: محاسبه نویز به دست آمده در تصویر، ساخت ویژگی‌ها و طبقه‌بند دودویی. با پیشرفت یادگیری عمیق، محققان به استفاده از روش های مبتنی بر یادگیری عمیق در نهان‌نگاری و نهان‌کاوی روی آوردند. با استفاده از یادگیری عمیق در نهان‌کاوی شکنی، مراحل استخراج ویژگی و طبقه‌بندی و با هم یکپارچه شده است و بدین ترتیب در این روش‌ها نیازی به استخراج ویژگی ها به صورت دستی نیست. در شکل ۱، ساختار کلی روش‌های نهان‌کاوی سنتی با استخراج دستی ویژگی‌ها (بخش بالا) و روش های مبتنی بر یادگیری عمیق (بخش پایین) نشان داده شده است.

در سال ۲۰۱۴، اولین ساختار CNN برای کشف روش های پنهان‌نگاری در حوزه مکان پیشنهاد شده است. اگرچه این مدل پیشنهادی بهتر از SPAM عمل می‌کند، اما نتوانست نسبت به SRM به عملکرد قابل قبولی برسد. این شبکه به خاطر وجود سه لایه پیچشی به اندازه کافی عمیق نیست و به دلیل وجود لایه بسیار بزرگ تمام متصل بسیار کند است [۱۸]. در ابتدای سال ۲۰۱۵، یک شبکه CNN برای پنهان‌شکنی طراحی شده است که عملکرد آن مشابه عملکرد روش دو مرحله‌ای (SRM+EC) است. در این مدل، با استفاده از فیلترها در شبکه توانستند محتوای تصویر را پنهان و نسبت سیگنال به نویز را بهبود دهند [۱۹].

یک سال بعد، مدلی پیشنهاد شد که در شرایط استفاده از کلید جاسازی یکسان برای تصاویر مختلف نسبت به مدل‌های قبلی موفق‌تر است، اما در صورت استفاده از کلیدهای متفاوت برای جاسازی در هر تصویر، عملکرد ضعیف‌تری دارد [۲۰]. در سال ۲۰۱۶، یک چارچوب مبتنی بر یادگیری عمیق، Xu-NetV1، که عملکردی قابل مقایسه با SRM داشت، ارائه شد. در این مدل از یک لایه ABS [1] و لایه‌های BN [2] و ادغام [3] استفاده شده است [۲۱]. در همان سال، یک نسخه بهبود یافته از این روش، Xu-NetV2، با هدف پیاده سازی یادگیری جمعی ارائه شد. در این مدل یک گروه از لایه‌ها به شبکه اضافه شد و به علاوه اندازه هسته ادغام نیز افزایش یافت [۲۲].

در سال ۲۰۱۷، مدل Ye-Net ارائه شد که با استفاده از تکنیک‌هایی مانند افزایش داده [4]، توانست عملکرد بهتری نسبت به SRM داشته باشد [۲۳]. Yedroudj-Net [۱۷]، ReST-Net [۲۴] و SRNet [۲۵] از بهترین شبکه‌های پیشنهاد شده در سال ۲۰۱۸ هستند. Yedroudj-Net شبکه کوچکی است که نیازی به مجموعه داده بزرگ ندارد و بدون استفاده از یادگیری انتقالی [5] و افزایش داده عملکرد مناسبی دارد. در مقابل، ReST-Net یک شبکه بزرگ است که از سه زیرشبکه ساخته شده است. SRNet به مجموعه داده بزرگتری نسبت به Yedroudj-Net نیاز دارد. در سال ۲۰۱۹، مدل Zhu-Net پیشنهاد شد که اندازه هسته لایه پیش پردازش را بهینه کرده است و در کشف روش های S-UNIWARD و WOW موفق بوده است [۲۶].

در یکی از مدل‌های ارائه شده در ۲۰۲۰، از الگوریتم k-means برای خوشه‌بندی داده‌ها استفاده شده است. هر خوشه داده به شبکه‌های CNN جداگانه داده شده و در مرحله آخر این شبکه‌ها ادغام شده‌اند [۲۷]. ایده مدل دیگر پیشنهاد شده در این سال، کاهش پیچیدگی از طریق گام پیش‌پردازش بود [۲۸].

از جدیدترین کارها در این حوزه می‌توان به [۲۹]، [۳۰] و [۳۱] اشاره کرد. استفاده از توابع فعال‌ساز مناسب در [۲۹] با بهبود عملکرد همراه بوده است. مدل پیشنهادی در [۳۱] از سه ماژول تشکیل شده است: ماژول استخراج نویز، ماژول تحلیل نویز و ماژول رده‌بندی. در این مدل از یک روش ادغام جدید استفاده شده است.

بررسی مقالات نهان‌کاوی مبتنی بر یادگیری عمیق موجود نشان می‌دهد که در اغلب آنها از تعداد محدودی از روش‌های نهان‌نگاری مکانی برای سنجش عملکرد استفاده شده است. در جدول ۱، خلاصه‌ای از ساختار تعدادی از روش‌های نهان‌کاوی حوزه مکان مبتنی بر CNN به همراه روش‌های نهان‌نگاری بررسی شده لیست شده است. WOW و S-UNIWARD دو روش رایج به عنوان معیار سنجش در میان مقالات مختلف است. بنابراین روش‌های مکانی بسیاری وجود دارد که در مورد نحوه عملکرد روش‌های مبتنی بر یادگیری عمیق بر روی آنها اطلاعاتی وجود ندارد. به همین دلیل در این مقاله سعی شده است روش‌های مکانی بیشتری مورد بررسی قرار گیرد.

۳ مدل پیشنهادی

مدل استفاده شده در این مقاله، مشابه با مدل Yedroudj-Net [۱۷] می‌باشد. تفاوت مدل پیشنهادی و مدل پایه، استفاده از یک لایه جدید در هرگروه در جهت به تاخیر انداختن پدیده بیش‌برازش است. معماری مدل پیشنهادی در شکل ۲ نشان داده شده است. این مدل، از سه قسمت پیش‌پردازش، ماژول پیچشی و طبقه‌بند تشکیل شده است که در لایه ماژول

[1] Absolute value activation (ABS)
[2] Batch Normalization (BN)
[3] Pooling
[4] Data augmentation
[5] Transfer learning

جدول ۱. خلاصه‌ای از روش‌های نهان‌کاوی حوزه مکان مبتنی بر CNN

الگوریتم‌ها	ماژول پیش‌پردازش	تابع فعال‌ساز	ادغام	روش‌های تست
Tan-Net [۱۸]	یک لایه پیچشی با ۴۰ هسته ۵ × ۵	Sigmoind	ادغام بیشینه	HUGO
Qian-Net [۱۹]	یک فیلتر بالاگذر از پیش تعریف‌شده	ReLU, Gaussian	ادغام میانگین	WOW/SUNIWARD/HUGO
Xu-Net [۲۱]	یک فیلتر بالاگذر از پیش تعریف‌شده	ReLU, TanH	ادغام میانگین	S-UNIWARD/HILL
Ye-Net [۲۳]	۳۰ فیلتر بالاگذر از SRM	TLU, ReLU	ادغام میانگین	WOW/S-UNIWARD/HILL
Yedroudj-Net [۱۷]	۳۰ فیلتر بالاگذر از SRM	TLU, ReLU	ادغام میانگین	WOW/S-UNIWARD
ResT-Net [۲۴]	چند گروه از فیلترهای بالاگذر	ReLU, Sigmoid, TanH	ادغام میانگین	S-UNIWARD/HILL/ CMD-HILL
SRNet [۲۵]	--	ReLU	ادغام میانگین	WOW/S-UNIWARD/HILL/MiPOD
Zhu-Net [؟]	۳۰ فیلتر بالاگذر از SRM	ReLU	ادغام میانگین	WOW/S-UNIWARD/HILL
GBRAS-Net [۲۹]	۳۰ فیلتر بالاگذر از SRM	ELU, TanH	ادغام میانگین	WOW/S-UNIWARD
SNMC-Net [۳۰]	۳۰ فیلتر بالاگذر از SRM و دو لایه پیچشی	ReLU	ادغام میانگین	WOW/S-UNIWARD/HILL
CCNet [۳۱]	هفت لایه پیچشی	ReLU, Sigmoid	ادغام پیچشی	WOW/SUNIWARD/HUGO

پیچشی ۵ بلوک و در هر بلوک، ۶ گام وجود دارد. تعداد گام‌ها در مدل پیشنهادی به دلیل استفاده از لایه حذف تصادفی [1] از مدل Yedroudj-Net [۱۷] بیشتر می‌باشد. در ادامه جزئیات هر لایه بیشتر شرح داده می‌شود.

۱.۳ پیش‌پردازش

در این بخش، یک تصویر خاکستری با ابعاد ۱ × ۲۵۶ × ۲۵۶ به عنوان ورودی به مدل فرستاده می‌شود. در این لایه از یک فیلتر بالاگذر برای عملیات کانولوشن استفاده می‌شود و این فیلتر در طول آموزش ثابت نگه داشته می‌شود. با انجام این عملیات، نویز گنجانده شده داخل تصویر تقویت شده و تأثیر محتوای آن بسیار کمتر می‌شود. این لایه می‌تواند یک مقداردهی اولیه خوب برای هدایت کل شبکه باشد و در مقایسه با مقداردهی تصادفی شبکه عملکرد بسیار بهتری دارد. در مدل پیشنهادی فقط از یک فیلتر SRM با ابعاد ۵ × ۵ با هدف شناسایی نواحی لبه تصویر استفاده شده است. این فیلتر و نمونه‌ای از عملکرد این بخش در شکل ۳ نشان داده شده است.

۲.۳ ماژول پیچشی

همانطور که اشاره شد این بخش از ۵ بلوک و هر بلوک از ۶ گام تشکیل شده است که در ادامه هر گام معرفی می‌شود. بخش پیچشی ویژگی‌های با ابعاد ۱ × ۲۵۶ × ۲۵۶ را به عنوان ورودی گرفته و در نهایت ویژگی‌هایی با ابعاد ۱۲۸ × ۱ × ۱ را به بخش طبقه‌بندی ارسال می‌کند. نمونه‌ای از شمای داخلی این ماژول در شکل ۴ نشان داده شده است.

۱.۲.۳ لایه پیچشی

در بلوک‌های ۱ و ۲ از هسته‌های پیچشی ۵ × ۵ و در بلوک‌های ۳ تا ۵، هسته‌های پیچشی با ابعاد ۳ × ۳ قرار داده شده است. به منظور کمک به مدل‌سازی آماری شبکه همانند مدل Res-Net [۳۲]، بایاس پیش فرض غیرفعال شده است تا ویژگی‌های تولید شده نسبت به صفر متقارن باشند. مقدار گام [2] و لایه‌گذاری [3] مورد استفاده در این لایه‌ها به این صورت تعیین می‌شود که همزمان با ادغام کردن ورودی و فیلترها کاهش اندازه هم انجام شود و نتیجه کار بعد از ادغام کوچک‌تر باشد.

۲.۲.۳ لایه ABS

این لایه مدل آماری را مجبور می‌کند تا تقارن موجود در نویزهای باقیمانده را درنظر بگیرد. تقارن به این معنی است که در صورت منفی کردن تصویر، ویژگی‌های آماری آن تغییر نمی‌کند. این لایه فقط در بلوک ۱ روش ۱llrXu-NetV [۲۱] استفاده شده است.

۳.۲.۳ لایه‌های BN و مقیاس

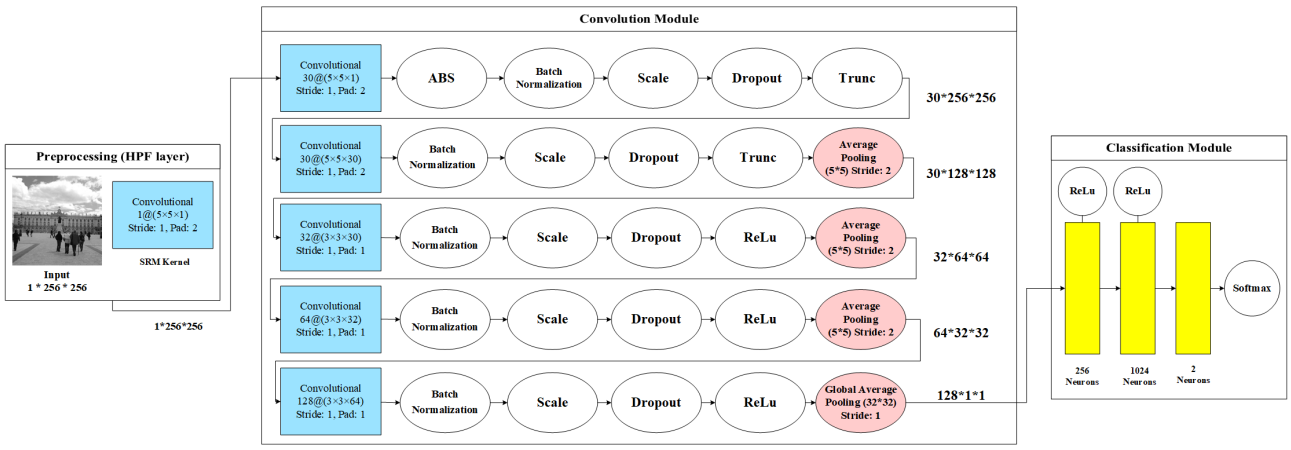
لایه BN هر یک از ویژگی‌ها را به گونه‌ای نرمالیزه می‌کند که توزیع ویژگی حاصل دارای میانگین صفر و انحراف معیار یک باشد و در نهایت این توزیع را مقیاس و جابجا می‌کند. با استفاده از یک لایه BN می‌توان از نرخ یادگیری بزرگتر برای افزایش سرعت یادگیری استفاده کرد و دقت تشخیص بهبود می‌یابد. در مدل پیشنهادی از لایه مقیاس [4] همراه با لایه BN استفاده شده است.

۴.۲.۳ لایه حذف تصادفی

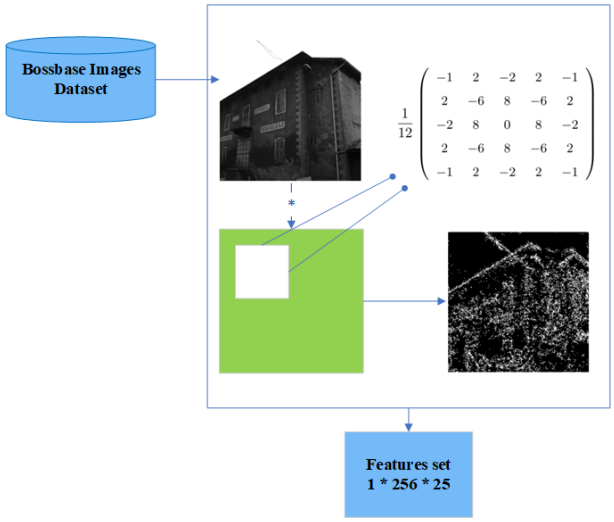
در یک شبکه عصبی که شامل تعدادی نورون است، لایه حذف تصادفی باعث می‌شود که شبکه در حین آموزش نورون‌ها، از تعدادی از آن‌ها به صورت تصادفی چشم‌پوشی کند. به عبارت دیگر، آن نورون‌های خاص، در مسیر رفت یا برگشت در نظر گرفته نمی‌شوند. در روش پیشنهادی از این لایه جهت به تاخیر انداختن پدیده بیش‌برازش در تعداد تکرارهای بالا استفاده شده است.

[1] Dropout

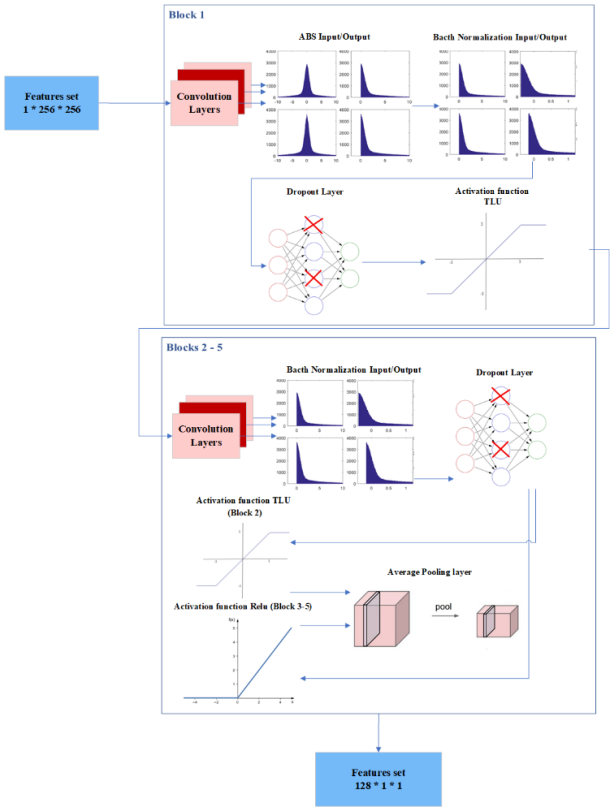
[2] Stride [3] Padding [4] Scale



شکل ۲. معماری مدل پیشنهادی



شکل ۳. مثالی از بلاک پیش‌پردازش



شکل ۴. مثالی از ماژول پیچشی

۵.۲.۳ لایه فعال‌ساز

تابع Trunc، نسخه‌ای از تابع ReLU است که طبق فرمول (۱) تعریف شده است. مقدار پارامتر T از طریق آزمایش‌های مختلف تعیین می‌شود.

$$Trunc(x) = \begin{cases} -T & x < -T \\ x & -T \le x \le T \\ T & x > T \end{cases} \qquad (۱)$$

در مدل پیشنهادی، از این تابع در بلوک‌های ۱ و ۲ استفاده شده است. استفاده از تابع Trunc در لایه‌های اولیه پیچشی به موارد زیر کمک می‌کند [۲۴]:

- با توزیع سیگنال‌های ضعیف داخل تصویر سازگار است.
- شبکه را به یادگیری مؤثرتر فیلترهای HPF وادار می‌کند.

در [۲۳] نشان داده شده است که با به کارگیری این تابع در لایه‌های عمیق‌تر، دقت شبکه کاهش می‌یابد، بنابراین برای بلوک‌های ۳ تا ۵، تابع ReLU به دلیل عملکرد خوب و محاسبات سریع، مورد استفاده قرار گرفته است.

۶.۲.۳ لایه ادغام میانگین

خروجی تابع فعال‌ساز از لایه کاهش اندازه عبور داده شده که در آن با کاهش اندازه خروجی، اطلاعات مهم حفظ و جزئیات نادیده گرفته می‌شوند. در

روش ادغام بیشینه، ممکن است برخی از اطلاعات مفید از دست بروند. به همین دلیل، از روش ادغام میانگین که عملکرد بهتری نسبت به روش اول دارد استفاده شده است. این لایه در بلوک‌های ۲ تا ۵ استفاده شده است. در آخرین بلوک، یک ادغام میانگین سراسری برای تولید یک به یک عناصر متناظر با نگاشت ویژگی مورد استفاده قرار می‌گیرد و برای جلوگیری از فقدان اطلاعات در ابتدای شبکه، هیچ کاهش بعدی در بلوک اول وجود ندارد.

۳.۳ ماژول طبقه‌بندی

بخش طبقه‌بند که از ۳ لایه تمام متصل تشکیل شده است. در مدل پیشنهادی، لایه اول شامل ۲۵۶ و لایه دوم ۱۰۲۴ نورون است که در این دو لایه از تابع فعال‌ساز ReLU استفاده شده است. آخرین لایه تمام متصل در این بخش حاوی دو نورون که متناظر با پوشانه و گنجانه بودن تصاویر است، می‌باشد. در نهایت برای تولید توزیع در دو برچسب کلاس از تابع فعال ساز softmax استفاده شده است.

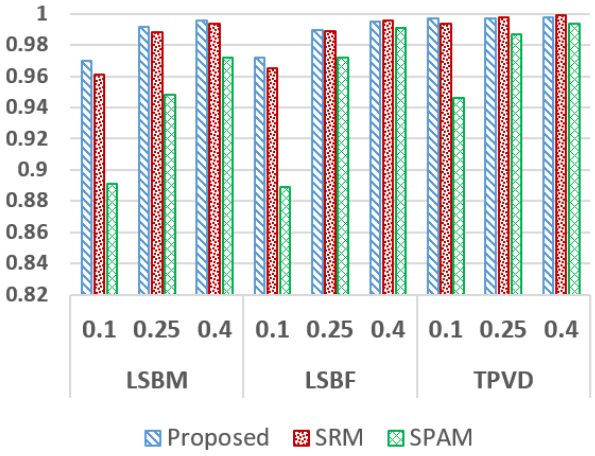
۴ نتایج پیاده سازی

مدل پیشنهادی با زبان پایتون و کتابخانه کراس پیاده‌سازی شده است. بستر سخت افزاری استفاده شده برای تست مدل پیشنهادی به شرح زیر است:

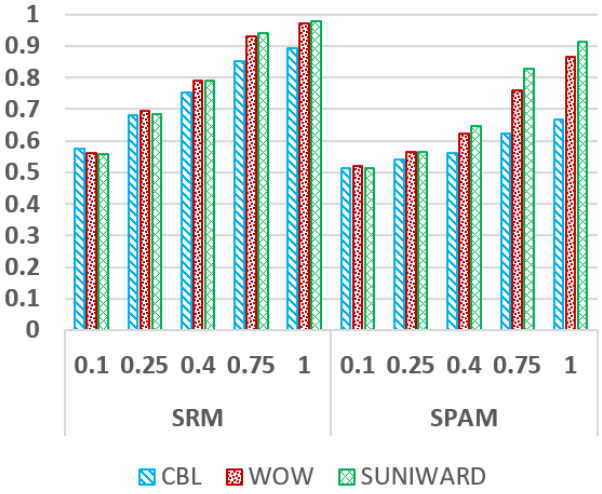
- پردازنده: Intel(R) Core(TM) i78700k CPU @ 3.70GHZ
- حافظه RAM: 32GB
- پردازنده گرافیکی: NVIDIA GeForce GTX 1080 Ti

مجموعه تصاویر BOSSbase که شامل ۱۰۰۰۰ تصویر خاکستری با ابعاد ۵۱۲ × ۵۱۲ می‌باشد، برای آموزش و تست شبکه مورد استفاده قرار گرفته است. برای کاهش بار محاسباتی و مطابقت با ابعاد ورودی شبکه پیشنهادی، ابتدا ابعاد این تصاویر به ۲۵۶ × ۲۵۶ کاهش می‌یابند. برخلاف روش Yedroudj-Net که بعد از حداکثر ۳۰۰ تکرار با مسئله بیش‌برازش روبرو شده است [۱۷]، تعداد تکرارها در آموزش مدل پیشنهادی، ۲۰۰۰ در نظر گرفته شده است. در این تست مقدار نرخ آموزشی برابر ۰٫۰۰۰۱، اندازه دسته تصاویر آموزش برابر ۳۲ و برای تصاویر اعتبارسنجی برابر ۶۴ تنظیم شده است.

ابتدا از مدل پیشنهادی برای کشف الگوریتم‌های نهان‌نگاری LSBM، LSBF و TPVD استفاده شد. مقدار دقت برای کشف این سه الگوریتم نهان‌نگاری در جاسازی‌های ۰٫۱، ۰٫۲۵ و ۰٫۴ با استفاده از مدل پیشنهادی و روش‌های سنتی در شکل ۵ نمایش داده شده است. بررسی این نمودار نشان می‌دهد که در درصد جاسازی‌های پایین موفقیت روش‌های سنتی کمتر از روش پیشنهادی است. با افزایش داده جاسازی شده، این روش‌ها راحت‌تر می‌توانند وجود داده را شناسایی کنند و حتی روش SRM+EC در مواردی عملکرد اندکی بهتر از روش پیشنهادی نیز دارد.



شکل ۵. دقت مدل پیشنهادی و روش‌های سنتی



شکل ۶. دقت روش‌های سنتی در کشف CBL و دو روش مکانی‌دیگر

برای تست مدل پیشنهادی در برابر روش‌های تطبیقی، روش CBL انتخاب شده است که به دلیل شناسایی نواحی لبه تصویر پوشانه و جاسازی داده در آنها، از امنیت خوبی در برابر حملات سنتی برخوردار است و حتی نسبت به روش‌های WOW و SUNIWARD نیز که به دلیل امنیت بالا، در اغلب مقالات به عنوان روش‌های تست استفاده می‌شوند، از امنیت بیشتری برخوردار است. برای اثبات این ادعا در شکل ۶، عملکرد دو روش سنتی SPAM+EC و SRM+EC برای کشف روش‌های WOW، SUNIWARD و CBL مقایسه شده است. نتایج ارائه شده در شکل ۶، درستی این ادعا را تأیید می‌کند.

در جدول ۲ مقدار دقت روش پیشنهادی و روش‌های سنتی در کشف CBL ارائه شده است. این نتایج نشان می‌دهد که مدل پیشنهادی در سطوح جاسازی مختلف توانسته با دقت بیشتر و خطای کمتر، روش CBL را کشف کند. بهبود رخ داده در سطوح جاسازی کم بسیار چشم‌گیر است. برای مثال، در جاسازی ۰٫۱، بهترین دقت روش‌های سنتی برابر ۰٫۵۷۸ و برای مدل پیشنهادی ۰٫۷۲۵ است. به علاوه دقت روش پیشنهادی در

جدول ۲. دقت مدل‌های مختلف در کشف روش CBL

	نرخ جاسازی (تعداد بیت در هر پیکسل)				
	۰٫۱	۰٫۲۵	۰٫۴	۰٫۷۵	۱
روش پیشنهادی	۰٫۷۲۵	۰٫۸۲۸	۰٫۸۷۷	۰٫۹۲۷	۰٫۹۳۹
Yedroudj-Net [۱۷]	۰٫۶۹۸	۰٫۸۰۵	۰٫۸۵۹	۰٫۹۱۱	۰٫۹۳۰
SRM+EC	۰٫۵۷۸	۰٫۶۸۱	۰٫۷۴۲	۰٫۸۵۲	۰٫۸۹۲
SPRM+EC	۰٫۵۱۵	۰٫۵۴۲	۰٫۵۶۲	۰٫۶۲۲	۰٫۶۶۶

کشف روش CBL در مقایسه با روش پایه یعنی Yedroudj-Net [۱۷] بهبود یافته است.

۵ نتیجه‌گیری

امروزه استفاده از شبکه‌های CNN برای شناسایی تصاویر حاوی داده پنهان یکی از حوزه‌های جذاب و موفق برای محققان است. تا به حال مدل‌های موفقی در این حوزه برای کشف روش‌های مکانی پیشنهاد شده است. در این مقاله یک معماری مبتنی بر CNN برای کشف روش‌های مکانی پیشنهادشده است و سعی شده است در مرحله تست، بررسی جامعی بر روی روش‌های مکانی مختلف انجام شود. مدل پیشنهادی، نسخه بهبود یافته مدل Yedroudj-Net است. یکی از تغییرات اعمال شده، اضافه کردن لایه حذف تصادفی با هدف به تاخیر انداختن پدیده بیش‌برازش است. با بررسی نتایج تست مدل پیشنهادی برای سه گروه از روش‌های مکانی براساس معیارهای ارزیابی مختلف مشخص شد که روش پیشنهادی با دقت بسیار بالا قادر به کشف روش‌های مبتنی بر LSB و PVD است. روش پیشنهادی در کشف روش CBL که یک روش تطبیقی است، نیز موفق است و این موفقیت در درصدهای جاسازی پایین نسبت به روش‌های سنتی بسیار چشمگیرتر است.

مراجع

[1] Li, Bin, Wang, Ming, Huang, Jiwu, and Li, Xiaolong. A new cost function for spatial image steganography. in *2014 IEEE International conference on image processing (ICIP)*, pp. 4206–4210. IEEE, 2014.

[2] Sedighi, Vahid, Cogranne, Rémi, and Fridrich, Jessica. Content-adaptive steganography by minimizing statistical detectability. *IEEE Transactions on Information Forensics and Security*, 11(2):221–234, 2015.

[3] Holub, Vojtěch, Fridrich, Jessica, and Denemark, Tomáš. Universal distortion function for steganography in an arbitrary domain. *EURASIP Journal on Information Security*, 2014:1–13, 2014.

[4] Holub, Vojtěch and Fridrich, Jessica. Designing steganographic distortion using directional filters. in *2012 IEEE International workshop on information forensics and security (WIFS)*, pp. 234–239. IEEE, 2012.

[5] Wu, Songtao, Zhong, Shenghua, and Liu, Yan. Deep residual learning for image steganalysis. *Multimedia tools and applications*, 77:10437–10453, 2018.

[6] Mandal, Pratap Chandra, Mukherjee, Imon, Paul, Goutam, and Chatterji, BN. Digital image steganography: A literature survey. *Information sciences*, 2022.

[7] Paul, Goutam, Saha, Sanjoy Kumar, and Burman, Debanjan. A pvd based high capacity steganography algorithm with embedding in non-sequential position. *Multimedia Tools and Applications*, 79(19-20):13449–13479, 2020.

[8] Sabeti, Vajiheh, Samavi, Shadrokh, and Shirani, Shahram. An adaptive lsb matching steganography based on octonary complexity measure. *Multimedia tools and applications*, 64:777–793, 2013.

[9] Pevnỳ, Tomáš, Filler, Tomáš, and Bas, Patrick. Using high-dimensional image models to perform highly undetectable steganography. in *Information Hiding: 12th International Conference, IH 2010, Calgary, AB, Canada, June 28-30, 2010, Revised Selected Papers 12*, pp. 161–177. Springer, 2010.

[10] Lyu, Siwei and Farid, Hany. Steganalysis using color wavelet statistics and one-class support vector machines. in *Security, steganography, and watermarking of multimedia contents VI*, vol. 5306, pp. 35–45. SPIE, 2004.

[11] Kodovsky, Jan, Fridrich, Jessica, and Holub, Vojtěch. Ensemble classifiers for steganalysis of digital media. *IEEE Transactions on information forensics and security*, 7(2):432–444, 2011.

[12] Pevnỳ, Tomáš, Bas, Patrick, and Fridrich, Jessica. Steganalysis by subtractive pixel adjacency matrix. in *Proceedings of the 11th ACM workshop on Multimedia and security*, pp. 75–84, 2009.

[13] Fridrich, Jessica and Kodovsky, Jan. Rich models for steganalysis of digital images. *IEEE Transactions on information Forensics and Security*, 7(3):868–882, 2012.

[14] Denemark, Tomas, Sedighi, Vahid, Holub, Vojtech, Cogranne, Rémi, and Fridrich, Jessica. Selection-channel-aware rich model for steganalysis of digital images. in *2014 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 48–53. IEEE,

Signal Processing Letters, 25(5):650–654, 2018.

[25] Boroumand, Mehdi, Chen, Mo, and Fridrich, Jessica. Deep residual network for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 14(5):1181–1193, 2018.

[26] Zhang, Ru, Zhu, Feng, Liu, Jianyi, and Liu, Gongshen. Depth-wise separable convolutions and multi-level pooling for an efficient spatial cnn-based steganalysis. *IEEE Transactions on Information Forensics and Security*, 15:1138–1150, 2019.

[27] Abazar, Tayebe, Masjedi, Peyman, and Taheri, Mohammad. An efficient ensemble of convolutional deep steganalysis based on clustering. in *2020 6th International Conference on Web Research (ICWR)*, pp. 260–264. IEEE, 2020.

[28] Mustafa, Eslam M, Elshafey, Mohamed A, and Fouad, Mohamed M. Enhancing cnn-based image steganalysis on gpus. *J. Inf. Hiding Multim. Signal Process.*, 11(3):138–150, 2020.

[29] Reinel, Tabares-Soto, Brayan, Arteaga-Arteaga Harold, Alejandro, Bravo-Ortiz Mario, Alejandro, Mora-Rubio, Daniel, Arias-Garzón, Alejandro, Alzate-Grisales Jesús, Buenaventura, Burbano-Jacome Alejandro, Simon, Orozco-Arias, Gustavo, Isaza, and Raúl, Ramos-Pollán. Gbras-net: a convolutional neural network architecture for spatial image steganalysis. *IEEE Access*, 9:14340–14350, 2021.

[30] Han, Xu and Zhang, Tao. Spatial steganalysis based on non-local block and multi-channel convolutional networks. *IEEE Access*, 10:87241–87253, 2022.

[31] Fu, Tong, Chen, Liquan, Fu, Zhangjie, Yu, Kunliang, and Wang, Yu. Ccnet: Cnn model with channel attention and convolutional pooling mechanism for spatial image steganalysis. *Journal of Visual Communication and Image Representation*, 88:103633, 2022.

[32] He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, and Sun, Jian. Deep residual learning for image recognition. in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

2014.

[15] Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[16] Farooq, Numrena and Selwal, Arvind. Image steganalysis using deep learning: a systematic review and open research challenges. *Journal of Ambient Intelligence and Humanized Computing*, 14(6):7761–7793, 2023.

[17] Yedroudj, Mehdi, Comby, Frédéric, and Chaumont, Marc. Yedroudj-net: An efficient cnn for spatial steganalysis. in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2092–2096. IEEE, 2018.

[18] Tan, Shunquan and Li, Bin. Stacked convolutional auto-encoders for steganalysis of digital images. in *Signal and information processing association annual summit and conference (APSIPA), 2014 Asia-Pacific*, pp. 1–4. IEEE, 2014.

[19] Qian, Yinlong, Dong, Jing, Wang, Wei, and Tan, Tieniu. Deep learning for steganalysis via convolutional neural networks. in *Media Watermarking, Security, and Forensics 2015*, vol. 9409, pp. 171–180. SPIE, 2015.

[20] Pibre, Lionel, Jérôme, Pasquet, Ienco, Dino, and Chaumont, Marc. Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover source-mismatch. *arXiv preprint arXiv:1511.04855*, 2015.

[21] Xu, Guanshuo, Wu, Han-Zhou, and Shi, Yun-Qing. Structural design of convolutional neural networks for steganalysis. *IEEE Signal Processing Letters*, 23(5):708–712, 2016.

[22] Xu, Guanshuo, Wu, Han-Zhou, and Shi, Yun Q. Ensemble of cnns for steganalysis: An empirical study. in *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*, pp. 103–107, 2016.

[23] Ye, Jian, Ni, Jiangqun, and Yi, Yang. Deep learning hierarchical representations for image steganalysis. *IEEE Transactions on Information Forensics and Security*, 12(11):2545–2557, 2017.

[24] Li, Bin, Wei, Weihang, Ferreira, Anselmo, and Tan, Shunquan. Rest-net: Diverse activation modules and parallel subnets-based cnn for spatial image steganalysis. *IEEE*

MonadiAfta-Bianual Scientific Journal for Cyberspace Security
2023, Volume 21, Number 2 (pp. 42–50)
https://monadi.isc.org.ir

Presented at the ISCISC 2023 in Iranian Research Organization for Science & Technology, Tehran, Iran

A comprehensive evaluation of deep learning based steganalysis performance in detecting spatial methods★

Vajiheh Sabeti * and Mahdiyeh Samiei

Department of Computer Engineering, Faculty of Engineerning, Alzahra University, Tehran, Iran

ARTICLE INFO.

Keywords:
Steganalysis
Spatial-based steganography
Deep learning
Convolutional neural network

dor: 20.1001.1.24763047.1402.12.2.5.1

Type: Research paper

ABSTRACT

Steganalysis is the art of detecting the existence of hidden data. Recent research has revealed that convolutional neural networks (CNNs) can detect data through automatic feature extraction. Several studies investigated the performance of existing models using a limited number of spatial steganography methods. This study aims to propose a CNN and comprehensively investigate its efficiency in detecting different spatial methods. The proposed model comprises three modules: preprocessing, convolutional (five blocks), and classifier (three fully connected layers). The test results for the least-significant-bit (LSB) and pixel-value differencing (PVD) based methods indicate that the proposed method can detect data of even concise length with high accuracy and a low error. The proposed method also detects complexity-based LSB-M (CBL) as an adaptive approach. Lower embedding rates make this success even more impressive. Manual feature extraction has much lower success rates due to low variations of statistical features at low embedding rates than the proposed model.

★ The ISCISC 2023 Program Committee effort is highly acknowledged for reviewing this paper.
* Corresponding author
Email addresses: v.sabeti@alzahra.ac.ir (Vajiheh Sabeti), samiee.mahdis@yahoo.com (Mahdiyeh Samiei)