



منطبق با محتوا

بهناز عبدالهی، احد هراتی*، امیر حسین طاهری‌نیا

گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران

چکیده

نهان‌نگاری هنر انتقال اطلاعات از طریق یک ارتباط محرمانه است. در نهان‌نگاری، اطلاعات حساس و مهم در یک محیط رسانه‌ای به نام پوشانه جاسازی می‌شود؛ به‌طوری‌که رسانه حاوی پیام از نمونه رسانه اصلی قابل تمایز نبوده و وجود پیام مخفی حتی به‌صورت احتمالی قابل تشخیص نباشد. اعوجاج حاصل از جاسازی در نهان‌نگاری منطبق با محتوا (تطبیقی) به ساختار محلی تصویر وابسته است؛ از این‌رو، تغییرات در مناطق پیچیده کمتر قابل تشخیص بوده و در نتیجه از اولویت بالاتری برای جاسازی برخوردار خواهند بود. تاکنون رویکردهای مختلفی در زمینه نهان‌نگاری منطبق با محتوا ارائه شده است: مبتنی بر مدل، مبتنی بر هزینه و مبتنی بر یادگیری تقابلی. در رویکرد نهان‌نگاری مبتنی بر مدل سعی می‌شود مدل آماری پوشانه تا حد ممکن حفظ شود؛ در حالی‌که هدف رویکرد مبتنی بر هزینه، کمینه‌سازی اعوجاج حاصل از مجموع هزینه‌های ویرایش پیکسل‌های حامل پیام است. در رویکرد یادگیری تقابلی، از رابطه رقابتی بین نهان‌نگار و نهان‌کاو برای حفظ مشخصات آماری تصویر و بهبود محرمانگی بهره گرفته می‌شود. در این مقاله مفاهیم و رویکردهای نهان‌نگاری معرفی می‌شود و سپس روش‌های پیشنهادی در نهان‌نگاری مورد بحث و بررسی قرار می‌گیرند.

واژه‌گان کلیدی: نهان‌نگاری منطبق با محتوا، نهان‌کاو، کمینه‌سازی اعوجاج، مدل آماری، یادگیری تقابلی.

A review of Content Adaptive Image Steganography methods

Behnaz Abdollahi, Ahad Harati* and Amir Hossein Taherinia

Computer Engineering Department, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract

Steganography is the art of transferring information through secret communication. The essential aim of steganography is to minimize the distortion caused by embedding the secret message; so that the image containing the message (stego) cannot be distinguished from the original image (cover), and the existence of the hidden message cannot be detected.

The distortion in content-adaptive steganography depends on the local structure of the image. The embedding changes into the areas with rich textures are less detectable than smooth areas, so the textured areas have a higher modification priority. In this regard, three main steganography approaches are proposed: model-based, cost-based, and adversarial. The model-based approach considers a statistical model for the cover image and tries to preserve this model during the embedding process. The cost-based one focuses on minimizing the distortion obtained from the sum of the heuristic costs of modified pixels. The adversarial approach uses the competition between steganography and steganalysis to improve the embedding performance.

In the first section of this paper, the concept of steganography and its history is expressed. Digital steganography including three types of cover synthesis, selection, and modification is introduced in the second section. The focus of this paper is on steganography based on the cover modification. The goal is to estimate the best probability distribution of modifications, and embedding the message in the estimated places is left to existing coding algorithms. In the third section, the problem of estimating the probability distribution is formulated as an optimization problem with the aim of distortion minimization. The distortion-based methods compute the probability distribution of embedding changes

* Corresponding author

* نویسنده عهده‌دار مکاتبات

using a pre-defined distortion function. In the additive distortion function, the embedding changes are assumed to be independent. Thus, the distortion function cannot capture interactions between changes caused by embedding, and it leads the performance to suboptimality. In this regard, the non-additive distortion functions are presented that consider the dependencies among the modification of adjacent pixels. The distortion-based methods include two model-based and cost-based approaches are introduced in the fourth section. Then, their most significant methods are reviewed in the fifth section.

Considering the competitive nature of steganography and steganalysis, a new steganography approach is presented in the sixth section that takes advantage of adversarial learning to improve secrecy. Adversarial learning includes two strategies: Generative adversarial networks (GANs) and adversarial attacks. In the concept of steganography, the GAN-based strategy tries to train the steganographic network against a steganalysis network. This is an iterative and dynamic game between steganographic and steganalysis networks to reach the Nash equilibrium. Another strategy attempts to simulate an adversarial attack and generate stego images that deceive the steganalysis network. The adversarial-based steganography methods are reviewed in the seventh section.

In the eighth section, different methods are compared from various points of view. The results of this study show that some techniques, such as smoothing the embedding changes, considering the interactions between the changes, using side-informed information, and exploring adversarial networks, can help to estimate the proper embedding probability map and improve performance and security. In the ninth section, suggestions are stated that can be considered for future research. Finally, the conclusion is expressed in the tenth section.

Keywords: Content-adaptive steganography, Steganalysis, Distortion minimization, Statistical model, Adversarial learnin

پنهان کردن محتوای پیام است، نهان نگاری با تلاش برای
پنهان کردن واقعیت ارتباطات، پا را فراتر می گذارد.
نهان نگاری و ته نقشن نگاری دو شاخه اصلی در
مخفی سازی اطلاعات هستند. در سامانه های مخفی سازی
اطلاعات سه شاخه ظرفیت^۵، محرمانگی^۶ و مقاومت^۷ حائز
اهمیت هستند. ته نقشن نگاری برای حفظ داده های
دیجیتال در مقابل رونوشت یا دست کاری های غیرمجاز
مورد استفاده قرار می گیرد. از این رو، مقاومت و پایداری
ته نقشن در برابر دست کاری یا حذف، اهمیت بسیاری
دارد. در برخی از کاربردها، حتی ممکن است جاسازی
داده و وجود یک ته نقشن دیجیتال در پوشانه، مسئله ای
آشکار و علنی بوده و اندازه ی نسبی ته نقشن هم بسیار
کوچک باشد. در صورتی که در نهان نگاری، امنیت و
غیرقابل شناسایی بودن اطلاعات با حفظ کیفیت پوشانه
در درجه نخست اهمیت و پس از آن هدف ظرفیت بالای
جاسازی است. در نهان نگاری تشخیص وجود پیام منجر
به شکست می شود؛ در حالی که در ته نقشن نگاری
مخفی بودن اطلاعات اهمیتی ندارد اما هدف پایداری و
حفظ اطلاعات جاسازی شده در برابر حملات و
آسیب های رسانه است.

ظهور هنر نهان سازی اطلاعات، بسیار قدیمی و
به طور تقریبی معاصر با رمزنگاری است. نخستین نمونه
مربوط به قرن پنجم قبل از میلاد است که در زندگی نامه
هرودت معرفی شده است [۱]. یونانی ها برای انتقال پیام

۱- مقدمه

امروزه، با توسعه اینترنت و پیشرفت در فناوری های
دیجیتال، ارتباطات نقشی اساسی در زندگی روزمره پیدا
کرده است؛ اما از طرف دیگر، انتشار اطلاعات خصوصی
می تواند خسارات زیادی را به دنبال داشته باشد. این امر
نگرانی هایی را در مورد امنیت اطلاعات ایجاد می کند.
محبوب ترین روش ها برای تأمین امنیت، رمزنگاری^۱ و
مخفی سازی داده^۲ است که مخفی سازی داده نیز به دو
دسته نهان نگاری^۳ (پنهان نگاری) و ته نقشن نگاری^۴ تقسیم
می شود.

روش های رمزنگاری پیام اصلی را به متن غیرقابل
فهم تبدیل می کنند، سپس متن رمزگذاری شده از
طریق یک کانال عمومی منتقل می شود؛ اما فقط شخص
مجاز که دارای یک کلید رمزگذاری است، می تواند آن را
رمزگشایی کند. با استفاده از روش های استاندارد
رمزگذاری می توان محتوای پیام را از دید ناظر غیرمجاز
مخفی نگه داشت؛ اما در بیشتر موارد، ارتباط با پیام های
رمزگذاری شده می تواند توجه ها را به خود جلب کند؛ زیرا
کانال ارتباطی می تواند کنترل شده و کلیه پیام های
عبوری از کانال ارتباطی تجزیه و تحلیل شوند. در چنین
حالتی، مؤثرترین راه عملی امنیت، پنهان کردن واقعیت
وجود ارتباط است که با کمک نهان نگاری قابل انجام
است. برخلاف روش های رمزگذاری که هدف آنها

¹ Cryptography

² Data hiding

³ Steganography

⁴ Watermarking

⁵ Capacity

⁶ Secrecy

⁷ Robustness

ارتباط مخفی هستند، نهان‌نگاری با شکست مواجه می‌شود.

با توجه به پیشرفت فناوری الکترونیک و دیجیتال، امروزه نهان‌نگاری دیجیتال بسیار موردتوجه قرار گرفته است که در آن یک پیام محرمانه با ابعاد به‌نسبه بزرگ در یک رسانه دیجیتال مانند متن، صوت، تصویر و یا فیلم مخفی می‌شوند. تصاویر، به‌دلیل افزونگی ذاتی و عدم توانایی چشم انسان در تشخیص تغییرات کوچک در مقادیر شدت پیکسل، می‌توانند به‌عنوان پوشانه مناسب برای انتقال اطلاعات محرمانه مورد استفاده قرار گیرند. علاوه بر ارتباطات مخفی، از نهان‌نگاری می‌توان در کارت‌های شناسایی کارکنان برای ذخیره اطلاعات مربوط به فرد، برای ذخیره اطلاعات شبکه در بسته‌های *TCP/IP*، جاسازی اطلاعات بیمار در تصاویر پزشکی و غیره استفاده کرد. روش‌های موجود بررسی‌شده در این مقاله منحصر به نهان‌نگاری در تصاویر دیجیتال هستند. پیام مخفی می‌تواند یک جریان بیت، متن ساده، متن رمزگذاری شده یا حتی یک تصویر باشد.

ادامه این مقاله به شرح زیر سازمان‌دهی شده است: در دو بخش بعدی تعاریف مورد نیاز در زمینه نهان‌نگاری بیان می‌شوند. در بخش چهارم، نهان‌نگاری مبتنی بر بهینه‌سازی تابع اعوجاج معرفی می‌شود که بر یافتن بستر مناسب برای جاسازی تمرکز دارد؛ به‌گونه‌ای که حداقل اعوجاج را در تصویر حامل ایجاد نماید و سپس در بخش پنجم به مرور روش‌های موجود در این حوزه پرداخته می‌شود. با توجه به رابطه رقابتی بین نهان‌نگاری و نهان‌کاوی، در بخش ششم، نحوه بهره‌گیری از یادگیری تقابلی در نهان‌نگاری بیان شده و سپس در بخش هفتم روش‌های مبتنی بر یادگیری تقابلی مورد بحث و بررسی قرار می‌گیرند. در بخش هشتم روش‌های مورد مطالعه مقایسه و ارزیابی می‌شوند. در بخش نهم، چشم‌انداز و مسیرهای آینده مشخص و در پایان نتیجه‌گیری نهایی بیان می‌شود.

۲- نهان‌نگاری

نهان‌نگاری به‌عنوان هنر انتقال اطلاعات از طریق یک ارتباط محرمانه شناخته شده است؛ به‌طوری‌که اطلاعات حساس و مهم در محیط رسانه‌ای به نام پوشانه^۱ جایگذاری می‌شود و حامل^۲ را تولید می‌کند شکل (۱). مهم‌ترین اصل در نهان‌نگاری، محرمانگی است؛ ازاین‌رو،

موهای یک برده را می‌تراشیدند، سپس پیامی را روی سر او خال‌کوبی می‌کردند. هنگامی که موها به عقب رانده می‌شد، برده می‌توانست بدون برانگیختن شک از خاک دشمن عبور کند. پس از حضور در محل گیرنده، پیام با تراشیدن سر برده دریافت می‌شد. مثال دیگر استفاده از جوهرهای نامرئی است [۲]. این روش پرکاربردترین روش نهان‌نگاری در طول قرن‌ها بوده است. مثال دیگر نوشتن پیام با استفاده از آب‌لیمو، شیر یا برخی مواد شیمیایی خاص در وسط متن‌های نوشته‌شده با جوهر است که باعث می‌شود پیام پس از خشک‌شدن کاغذ ناپدید شود؛ اما یک شعله ساده یا آغشته‌کردن کاغذ به یک ماده شیمیایی پیام را آشکار خواهد ساخت.

اغلب روش‌های نهان‌نگاری بین قرن سیزدهم تا شانزدهم شامل متن نوشتاری بوده‌اند. به‌عنوان مثال، برخی از کاغذ سوراخ‌دار به‌عنوان ماسک بین فرستنده و گیرنده استفاده می‌کردند. با قراردادن ماسک بر روی متن، پیام مشخص می‌شد. فرانسیس بیکن [۳] دریافت که دو فونت متفاوت برای هر حرف می‌تواند برای جاسازی نمایش دودویی پیام‌ها اعمال شود. با توجه به وضعیت نویسه‌نگاری در آن زمان، این روش خیلی قابل توجه نبود. بروستر [۴] در سال ۱۸۵۷ یک فن بسیار ابتکاری ابداع کرد که بعدها در چندین جنگ مورد استفاده قرار گرفت. وی پیشنهاد کرد که پیام تا حدی شبیه ذرات گردوغبار کوچک شود؛ اما در صورت بزرگ‌نمایی همچنان قابل خواندن باشد. پس از اینکه فناوری به‌درستی توسعه یافت، آلمانی‌ها از این "ریزنقطه‌ها" در جنگ جهانی اول استفاده کردند.

یکی از مشهورترین تعاریف سامانه نهان‌نگاری "مسئله زندانیان" است که در سال ۱۹۸۳ توسط سیمونز ارائه شد [۵]. در این تعریف سنتی، آلیس و باب به جرم جنایی دستگیر و در دو سلول جداگانه قرار می‌گیرند. آن‌ها به فکر طراحی نقشه فرار هستند، زندانیان مجاز به ارسال پیام هستند؛ اما متأسفانه تمام ارتباطات بین آن‌ها توسط یک نگهبان به نام حوا کنترل می‌شود. در چنین شرایطی، آلیس و باب برای برقراری ارتباط بدون جلب توجه حوا، نیاز به برقراری ارتباط محرمانه دارند. زندانیان توافق کردند که از نهان‌نگاری برای این ارتباط استفاده کنند. آن‌ها پیش از دستگیری یک الگوریتم جاسازی و استخراج پیام طراحی کرده و در مورد یک کلید خصوصی به توافق رسیده‌اند. رویکرد حوا برای آزمون آماری جهت تشخیص پیام مخفی، نهان‌کاوی نامیده می‌شود. در این سناریو، هنگامی که حوا بفهمد که آلیس و باب در حال تبادل پیام توسط یک

^۱ Cover
^۲ Stego

تلاش‌های متعدد جهت یافتن پوشانه مناسب است که این روش را تاحدودی غیرعملی کرده است.

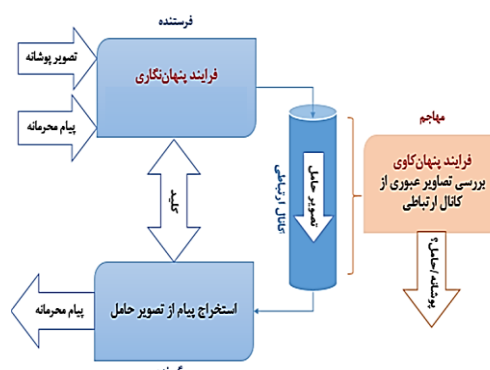
- در نهان‌نگاری با ساخت پوشانه، فرستنده به‌جای انتخاب پوشانه از یک پایگاه داده موجود، به‌سادگی یک پوشانه مناسب را ایجاد می‌کند که پیام را به بهترین شکل منتقل کند. اگر فرستنده بتواند پوشانه‌ای با توزیع مشخص بین خود و گیرنده تولید کند، می‌تواند پیام محرمانه را با آسودگی خیال پنهان کند. این روش از سطح امنیتی بالایی دارد، اما از نظر ظرفیت جاسازی بسیار محدود است که باعث محبوبیت کمتر این روش می‌شود.

- نهان‌نگاری با ویرایش پوشانه پرکاربردترین روش نهان‌نگاری است که مبتنی بر ویرایش یک پوشانه موجود برای جاسازی یک پیام محرمانه است؛ درحالی‌که تا حد امکان سعی می‌شود مؤلفه‌های آماری پوشانه حفظ شود. از آنجاکه نهان‌نگاری با ویرایش پوشانه، ویژگی‌های آماری پوشانه را در حین جاسازی پیام تغییر می‌دهد؛ بنابراین لازم است مناطق امن پوشانه به‌درستی انتخاب شوند تا در صورت ویرایش تأثیر قابل‌توجهی بر روی آماره‌های پوشانه نداشته باشند. این مهم در اغلب روش‌های نهان‌نگاری موجود با استفاده از یک تابع اعوجاج حاصل می‌شود که تأثیر جاسازی بر امنیت را مدل می‌کند.

در نهان‌نگاری با ویرایش پوشانه، اعوجاج ناشی از جاسازی پیام به محتوای تصویر وابسته است و این امر باعث تخمین احتمال یا نرخ تغییر بیشتر در نواحی پیچیده تصویر نسبت به نواحی هموار می‌شود؛ زیرا تشخیص الگوهای غیرمنطقی (حاصل از جاسازی) در نواحی پیچیده برای الگوریتم‌های نهان‌کاوی دشوارتر است. از این‌رو، نهان‌نگاری تطبیقی یا منطبق با محتوا^۵ مورد توجه پژوهش‌گران قرار گرفته است. در [۷] نهان‌نگاری غیر تطبیقی، ظرفیت جاسازی هر پیکسل ثابت است و بدون توجه به همسایگی اطراف تعیین می‌شود؛ بنابراین، اطلاعات در کل تصویر توزیع شده و تفاوتی بین نواحی مختلف تصویر وجود ندارد؛ درحالی‌که در نهان‌نگاری تطبیقی، ظرفیت جاسازی بر اساس ساختار محلی تصویر تعیین و هر پیکسل بر اساس احتمال به‌دست‌آمده از پیکسل‌های همسایگی محلی ویرایش می‌شود؛ از این‌رو، تغییرات ناشی از جاسازی^۶ در مناطقی که کمتر قابل تشخیص هستند، مانند بافت‌های غنی و مناطق نوفه‌ای، با اولویت بیشتری انجام می‌گیرد.

جایگذاری پیام باید به‌گونه‌ای انجام شود که مشخصات آماری پوشانه تا حد ممکن حفظ شود تا از حامل قابل تمایز نبوده و تشخیص موجودیت پیام ممکن نباشد؛ اما از طرف دیگر، ظرفیت جاسازی تا حد ممکن افزایش یابد. برخلاف تلاش روش‌های نهان‌نگاری برای مخفی‌سازی وجود پیام، مهاجمان پیام‌هایی را که از طریق کانال ارتباطی عبور می‌کنند با هدف تغییر، تخریب یا جلوگیری از انتقال تجزیه و تحلیل می‌کنند. زمینه پژوهشی که به بررسی روش‌های مقابله با نهان‌نگاری می‌پردازد، نهان‌کاوی^۱ نامیده می‌شود. نهان‌کاو لزوماً نباید پیام مخفی را استخراج کند؛ زیرا چنین کاری به‌طور معمول بسیار دشوار و یا غیرممکن است. در عوض، هدف نهان‌کاو این است که فقط وجود ارتباط محرمانه را با احتمال قابل قبولی تشخیص دهد [۶].

روش‌های نهان‌نگاری بر اساس چگونگی جاسازی پیام محرمانه به سه دسته تقسیم می‌شوند: نهان‌نگاری با انتخاب پوشانه^۲، نهان‌نگاری با ساخت پوشانه^۳، نهان‌نگاری با ویرایش پوشانه^۴.

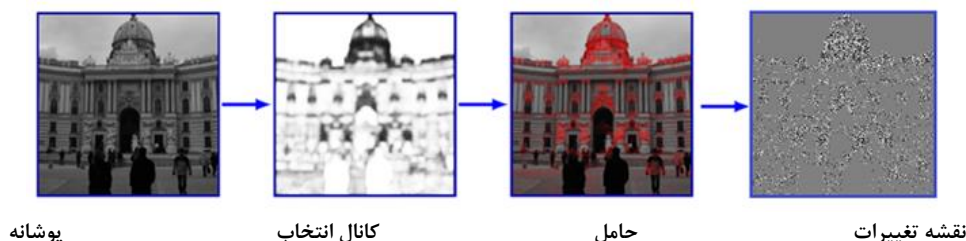


(شکل - ۱): فرایند نهان‌نگاری
Figure 1: The process of steganography

- در نهان‌نگاری با انتخاب پوشانه، فرستنده یک پایگاه داده از تصاویر در اختیار دارد. فرستنده پایگاه داده خود را با هدف یافتن تصویری متناسب با پیام موردنظر بررسی می‌کند. پس از یافتن و انتقال تصویر به همراه یک تابع درهم‌سازی و کلید مخفی، گیرنده می‌تواند با استفاده مجدد از تابع درهم‌سازی و کلید مشترک، به‌راحتی پیام مخفی را دریافت کند. از آنجاکه پوشانه در این روش تحت هیچ تغییری قرار نمی‌گیرد، از این‌رو غیرقابل‌شناسایی است. با این‌وجود، مشکل اصلی این روش‌ها ظرفیت جاسازی بسیار محدود و نیاز به

¹ Steganalysis
² Cover selection
³ Cover synthesis
⁴ Cover modification

⁵ Content-adaptive steganography
⁶ Embedding modifications



(شکل - ۲): مراحل نهان‌نگاری در حوزه مکان
Figure 2: The phases of spatial steganography

از این رو، مسئله محاسبه بستر جاسازی به صورت یک مسئله بهینه‌سازی فرموله می‌شود که هدف آن به کمینه‌رساندن اعوجاج^۴ کل است که از مجموع هزینه‌های مؤلفه‌های ویرایش شده بدست می‌آید [۲۵-۴۳] (شکل ۲).

این روش‌ها بر انتخاب پیکسل‌های مناسب برای جاسازی تأکید دارند و جاسازی پیام را به کدهای عملی کارآمد موجود می‌سپارند که با کمترین اعوجاج پیام را جاسازی می‌کنند.

الگوریتم جاسازی پیام: الگوریتم‌های جاسازی با ارائه یک روش کدگذاری مناسب، پیام را در پیکسل‌های پوشانه جایگذاری می‌کنند؛ به طوری که تا حد امکان اعوجاج کاهش و بازدهی جاسازی^۵ افزایش یابد. به عنوان مثال، روش جایگزینی بیت‌های کم‌ارزش که رایج‌ترین روش جاسازی پیام در حوزه مکان است و یا روش جاسازی ماتریسی^۶ با در نظر گرفتن هزینه جاسازی ثابت برای همه پیکسل‌های پوشانه، با بهره‌گیری از یک ساختار مناسب برای جاسازی پیام باعث کاهش تعداد تغییرات جاسازی در تصویر پوشانه و در نتیجه افزایش بازده جاسازی می‌شود [۸-۲۴].

این مقاله به مطالعه روش‌های نهان‌نگاری تطبیقی که بر انتخاب بستر جاسازی مناسب در حوزه مکان تمرکز دارند، می‌پردازد.

۳-۱- مؤلفه‌های سامانه نهان‌نگاری

سامانه نهان‌نگاری از یک منبع پوشانه، یک منبع پیام^۷ و توابع جاسازی^۸ و استخراج^۹ تشکیل شده است (شکل ۳). منبع پوشانه $\{C, P_c\}$ با مجموعه تمام تصاویر ممکن روی پوشانه $x \in C$ و توزیع آن‌ها تعریف می‌شود.

⁴ Distortion

⁵ Embedding efficiency

⁶ Matrix Embedding

⁷ Message

⁸ Embedding

⁹ Extraction

۳- نهان‌نگاری بر اساس ویرایش پوشانه

محرمانگی و غیرقابل شناسایی بودن نیاز اصلی رویکرد نهان‌نگاری است. قابلیت شناسایی داده‌های پنهان در یک شیء حامل، تحت تأثیر سه عنصر اساسی است:

- نوع رسانه پوشانه که حوزه جاسازی را مشخص می‌کند؛
- انتخاب مکان‌های مناسب جاسازی در پوشانه (بستر جاسازی یا کانال انتخاب^۱)؛
- الگوریتم جاسازی پیام.

پژوهش‌گران در حوزه نهان‌نگاری، به طور معمول تمرکز خود را بر روی یکی از این جنبه‌ها معطوف می‌سازند.

حوزه جاسازی: نهان‌نگاری را می‌توان بر اساس حوزه‌ای که پیام در آن جاسازی می‌شود به دو دسته مکان [۸-۴۳] و تبدیل [۴۴-۵۱] دسته‌بندی نمود. در نهان‌نگاری حوزه مکان^۲، پیام به طور مستقیم در پیکسل‌های تصویر پوشانه جاسازی می‌شود، در حالی که در نهان‌نگاری حوزه تبدیل^۳، ابتدا تصویر با یک تبدیل خاص (مانند DDT, Wavelet, FFT) به حوزه فرکانس برده شده و سپس پیام در ضرایب تبدیل جایگذاری می‌شود. روش‌های حوزه مکان به طور معمول از ظرفیت و کیفیت بیشتری برخوردارند؛ در حالی که به دلیل تغییر مستقیم پیکسل‌ها امکان شناسایی آن‌ها نیز بیشتر است؛ از این رو، در این روش‌ها تمرکز ویژه‌ای بر روی انتخاب مکان‌های مناسب جاسازی با قابلیت تشخیص کمتر وجود دارد.

بستر جاسازی: برای انتخاب مکان‌های مناسب جاسازی، به هر پیکسل از تصویر پوشانه یک مقدار عددی (معیار قابلیت تشخیص) اختصاص می‌یابد که تأثیر تغییر حاصل از جاسازی در آن پیکسل را اندازه‌گیری می‌کند و هزینه جاسازی نامیده می‌شود.

¹ Selection channel

² Spatial domain

³ Transforms domain

$$D_{KL}(P_C \parallel P_S) = \sum_{x \in \mathcal{C}} P_C(x) \log \frac{P_C(x)}{P_S(x)} \quad (1)$$

اگر $D_{KL}(P_C \parallel P_S) = 0$ باشد، به معنی یکسان بودن توزیع‌هاست $P_C = P_S$ و در این صورت نمی‌توان بین پوشانه و حامل تفاوت قائل شد و سامانه نهان‌نگاری به‌طور کامل امن خواهد بود. این امر بسیار مطلوب است؛ اما در واقعیت به‌راحتی نمی‌توان به آن دست یافت؛ بنابراین، یک سامانه نهان‌نگاری در صورتی امن نامیده می‌شود که تفاوت بین دو توزیع حامل و پوشانه تا حد ممکن ناچیز باشد $D_{KL}(P_C \parallel P_S) \leq \epsilon$.

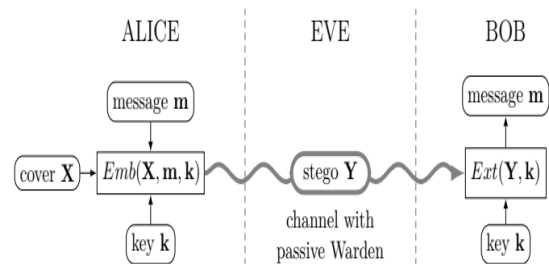
۲-۳- رویکردهای نهان‌نگاری

تاکنون سه رویکرد اصلی در زمینه نهان‌نگاری تطبیقی یا منطبق با محتوا ارائه شده است: رویکرد مبتنی بر مدل که بر پایه حفظ مدل آماری^۳ پوشانه عمل می‌کند، رویکرد مبتنی بر هزینه که هدف آن به کمینه‌رساندن اعوجاج^۴ حاصل از مجموع هزینه مؤلفه‌های تغییر یافته پوشانه است و رویکرد تقابلی^۵ که از محیط رقابتی بین نهان‌نگار و نهان‌کاو برای بهبود عملکرد جاسازی بهره می‌گیرد. دو رویکرد اول نهان‌نگاری سنتی مبتنی بر مفاهیم بهینه‌سازی هستند، درحالی‌که روش‌های رویکرد سوم بر اساس یادگیری عمیق^۶ طراحی شده‌اند.

- رویکرد نخست مبتنی بر حفظ مدل آماری پوشانه است که در آن طراح مدلی برای پوشانه در نظر می‌گیرد و جاسازی با هدف حفظ کامل و یا تقریبی مدل انجام می‌شود. در این رویکرد که بر اساس مشخصات آماری ارزیابی می‌شود، باید خصوصیات آماری پوشانه و حامل تا حد امکان یکسان باشد و اعوجاج حاصل از جاسازی به کمینه برسد. نهان‌نگاری با این روش بسیار ایمن خواهد بود؛ اما یافتن مدل مناسب دشوار است؛ زیرا یک مدل ساده قادر به پوشش‌دادن جزییات نبوده و از سوی دیگر مدل‌های پیچیده در حل با مشکل مواجه خواهد شد؛ ازاین‌رو، عدم تطبیق مدل مناسب می‌تواند توسط مهاجمان برای کشف اطلاعات پنهان مورد سوءاستفاده قرار گیرد.

- محور اصلی رویکرد دوم، به کمینه‌رساندن اعوجاج کل است. اعوجاج در این رویکرد از مجموع هزینه‌های اکتشافی مؤلفه‌های ویرایش‌شده پوشانه به‌دست

منبع پیام $\{\mathcal{M}, P_m\}$ به‌طور مشابه با مجموعه تمام پیام‌های ممکن و توزیع آن‌ها بیان می‌شود. توجه داشته باشید که توزیع P_m تابعی از منبع پوشانه است. با در نظر گرفتن $\mathcal{S}$ به‌عنوان مجموعه همه تصاویر حامل ممکن با توزیع P_s ، تابع جاسازی $\mathcal{S}: \mathcal{C} \times \mathcal{M} \times \mathcal{K} \rightarrow \mathcal{S}$ پیام را با استفاده از یک کلید مخفی مشترک $k \in \mathcal{K}$ در پوشانه جاسازی می‌کند؛ سپس تصاویر حامل $y = \text{Emb}(x, k, m) \in \mathcal{S}$ تولید شده از طریق یک کانال عمومی به گیرنده منتقل می‌شود. تابع استخراج $\text{Ext}: \mathcal{S} \times \mathcal{K} \rightarrow \mathcal{M}$ می‌تواند پیام مخفی $m = \text{Ext}(y, k)$ را با استفاده از کلید مشترک صحیح k استخراج کند. با داشتن منبع پوشانه $\{C, P_C\}$ ، هر شیء پوشانه را می‌توان به‌عنوان یک متغیر تصادفی در نظر گرفت که از توزیع پوشانه $X \sim P_C$ پیروی می‌کند. به‌طور مشابه، حامل نیز یک متغیر تصادفی با تابع توزیع حامل $Y \sim P_S$ در نظر گرفته می‌شود.



(شکل - ۳): مؤلفه‌های کانال نهان‌نگاری [۷]

Figure 3: Steganography Channel Components [7]

در رویکرد نهان‌نگاری، فرستنده و گیرنده به دنبال تبادل پنهانی اطلاعات هستند؛ درحالی‌که ناظر هر پیام را که از طریق کانال ارتباطی عبور کرده بررسی می‌کند و سوءظن به یک ارتباط مخفی منجر به قطع فوری آن کانال خواهد شد. ازاین‌رو، جاسازی باید به‌گونه‌ای انجام شود که ناظر از وجود این ارتباط مطلع نشود. بر اساس اصل کرشهف^۱، فرض می‌شود که ناظر به‌جز کلید مخفی، از همه چیز در مورد کانال نهان‌نگاری شامل منبع پوشانه، منبع پیام و همچنین توابع جاسازی و استخراج آگاه است.

اگر توزیع پوشانه و حامل از نظر آماری قابل تمایز نباشند، سامانه نهان‌نگاری مورد استفاده امن خواهد بود. یک معیار مناسب برای اندازه‌گیری فاصله بین این دو توزیع احتمال، واگرایی کولبک-لیبلر^۲ (KL) است که آنتروپی نسبی نیز نامیده می‌شود:

³ Model preservation

⁴ Distortion minimization

⁵ Adversarial

⁶ Deep learning

¹ Kerckhoffs' principle

² Kullback-Leibler (KL) divergence

است، طراحی تابع اعوجاج کاری چالش‌برانگیز است. تاکنون اعوجاج به دو صورت تعریف شده است: جمع‌شونده^۴ و غیرجمع‌شونده^۵. در تابع اعوجاج جمع‌شونده جاسازی اطلاعات در هر پیکسل مستقل از پیکسل‌های مجاور در نظر گرفته می‌شود؛ درحالی‌که در واقعیت تغییرات پیکسل‌ها به یکدیگر وابسته هستند و در تابع اعوجاج غیرجمع‌شونده سعی می‌شود اثر متقابل پیکسل‌ها لحاظ شود.

۴-۱- به‌کمینه‌رساندن اعوجاج جمع‌شونده

در این مقاله، پوشانه یک تصویر خاکستری در ابعاد $n_1 \times n_2$ است که با $x = \{x_1, x_2, \dots, x_n\} \in X$ ، $n = n_1 \times n_2$ نشان داده می‌شود و عملیات جاسازی، پیام را با ویرایش بیت‌هایی از برخی پیکسل‌ها در تصویر پوشانه جایگذاری می‌کند که منجر به تولید تصویر حامل $y = \{y_1, y_2, \dots, y_n\} \in Y$ می‌شود. پیکسل‌های پوشانه و حامل $((x_i, y_i), i = 1, 2, \dots, n)$ دارای مقدار عدد صحیح در محدوده $\{0, 255\}$ می‌باشند. طرح جاسازی نهان‌نگاری با نگاشتی که به جفت $\{Y, \pi\}$ اختصاص داده شده مرتبط است. $Y \subset X$ مجموعه تمام تصاویر حاملی y است که می‌تواند با جاسازی پیام در تصاویر پوشانه x تولید شود. هدف یافتن بستر جاسازی مناسب است که با تخمین تابع احتمال جاسازی بهینه $\pi(y)$ بدست می‌آید. این تابع احتمال تغییر پیکسل‌های پوشانه بر اثر مخفی‌سازی پیام را مشخص می‌کند $\pi(y) = P(Y = y)$.

روش تطبیق بیت با کمترین اهمیت (LSBM^۶) به‌عنوان یک الگوریتم جاسازی سه‌گانه، پیام را در تصویر پوشانه جاسازی می‌کند به‌طوری‌که $x_i - y_i \in I = \{-1, 0, +1\}$. درحالی‌که مقادیر $+1$ و -1 به ترتیب نشان‌دهنده افزایش و کاهش یک واحدی و صفر به معنی عدم تغییر مقدار پیکسل است؛ بنابراین، جاسازی به سه هزینه ویرایش برای هر پیکسل نیاز دارد $\{\rho_i^-, \rho_i^0, \rho_i^+\}$ که به ترتیب بیانگر هزینه کاهش، عدم تغییر و افزایش پیکسل پوشش‌دهنده است. ظرفیت نسبی جاسازی^۷ نیز بر اساس تعداد بیت‌های تغییر یافته در پیکسل^۸ محاسبه می‌شود.

می‌آید و در این روش سعی می‌شود حامل کمینه اعوجاج را نسبت به پوشانه دارا باشد. این روش منعطف‌تر بوده و امکان توسعه روش‌های نهان‌نگاری را که توسط عملکرد تشخیص نهان‌کاوی هدایت می‌شوند، فراهم می‌کند؛ ازاین‌رو، از سطح امنیتی بهتری برخوردار و رویکردی عملی‌تر است. این رویکرد، برخلاف ماهیت اکتشافی، در بسیاری از روش‌های نهان‌نگاری با امنیت مناسب مورد استفاده قرار گرفته است؛ اما ارتباط رسمی بین اعوجاج و تشخیص آماری وجود ندارد.

• کارهای اخیر نهان‌نگاری توانمندی مناسبی در استفاده از مفاهیم شبکه‌های عصبی مانند نمونه‌های تقابلی^۱ و شبکه‌های مولد تقابلی^۲ نشان داده‌اند. با توجه به حساست شبکه‌های عصبی به اغتشاشات کوچک تقابلی، در رویکرد سوم می‌توان از یادگیری تقابلی برای جاسازی مفهومی‌تر پیام بهره گرفت تا مشخصات آماری پوشانه تا حد ممکن حفظ شود. با توجه به رابطه رقابتی نهان‌نگاری و نهان‌کاوی، روش‌های نهان‌نگاری به‌عنوان یک حمله تقابلی^۳ سعی می‌کنند شبکه نهان‌کاو را گمراه کنند تا همه تصاویر ورودی را به‌عنوان تصاویر پوشانه تشخیص دهد. پس در مفهوم نهان‌نگاری، نهان‌کاو به‌عنوان تمایزدهنده در نظر گرفته می‌شود و شبکه‌های نهان‌نگار و نهان‌کاو به‌طور متناوب وزن‌های خود را با توجه به دانش به‌روز شده دیگری اصلاح می‌کنند.

۴- نهان‌نگاری مبتنی بر کمینه‌سازی اعوجاج

الگوریتم‌های نهان‌نگاری تطبیقی با هدف جاسازی پیام محرمانه در یک پوشانه، تلاش می‌کنند تا اعوجاج ناشی از روند جاسازی را به کمینه برسانند نمودار (۱). اعوجاج معیاری است که توسط یک نهان‌نگار برای مدل‌سازی اثر جاسازی پیام استفاده می‌شود و این روش‌ها سعی می‌کنند پیام را به‌گونه‌ای در پوشانه مخفی کنند که مشخصات آماری پوشانه تا حد ممکن حفظ شود.

هدف تعریف تابع اعوجاج، ارزیابی تأثیر آماری تغییرات حاصل از نهان‌نگاری است. از آنجایی‌که یک تصویر دیجیتال دارای ابعاد بالا و همبستگی‌های پیچیده

^۴ Additive Distortion

^۵ None-Additive Distortion

^۶ Least Significant Bit Matching

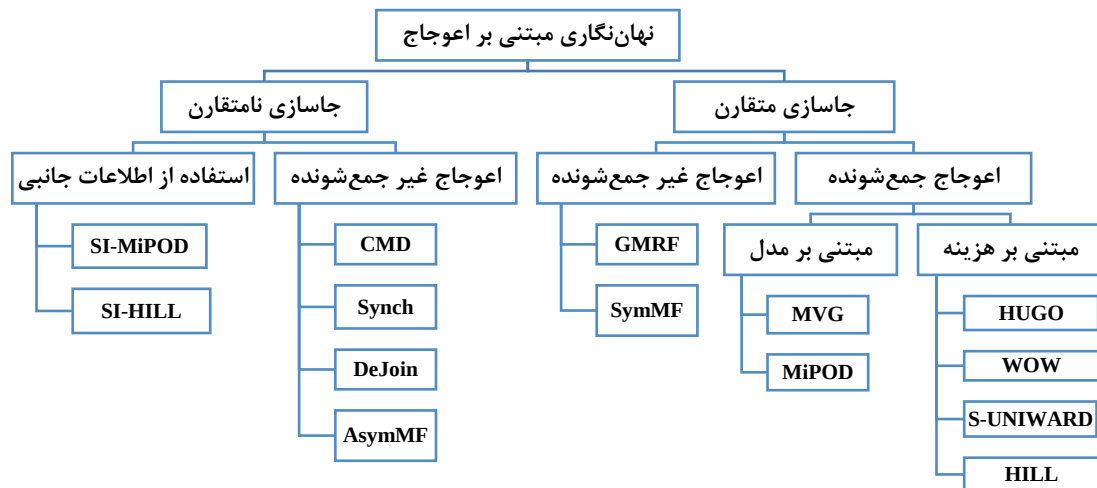
^۷ Relative payload

^۸ bits per pixel (bpp)

^۱ Adversarial examples

^۲ Generative adversarial networks

^۳ Adversarial attack



(نمودار - ۱) دسته‌بندی روش‌های نهان‌نگاری سنتی مبتنی بر کمینه‌سازی اعوجاج
Chart 1- Classification of traditional adaptive steganography methods based on distortion minimization

۴-۱-۱- رویکرد مبتنی بر هزینه

الگوریتم‌های نهان‌نگاری، با هدف جاسازی پیام محرمانه در یک پوشانه، تلاش می‌کنند تأثیر ناشی از روند جاسازی را به کمینه برسانند. گام نخست برای دستیابی به این هدف، معرفی یک معیار اعوجاج مناسب است. در این راستا، ابتدا یک هزینه به هر پیکسل پوشانه منتسب می‌شود که این هزینه منعکس‌کننده قابلیت تشخیص هر پیکسل در صورت ویرایش است. مجموع هزینه مؤلفه‌های تغییر یافته، میزان اعوجاج حامل را مشخص می‌کند؛ از این‌رو، هزینه کمتر بیانگر اولویت بیشتر مؤلفه برای جاسازی در راستای به کمینه‌رساندن اعوجاج پوشانه است؛ سپس پیام محرمانه در پوشانه بر اساس نقشه هزینه به دست‌آمده جاسازی می‌شود. فرایند جاسازی شامل ویرایش مؤلفه‌های پوشانه با کمک یک الگوریتم کدگذاری مناسب است؛ به‌طوری‌که کمترین اعوجاج ممکن در تصویر پوشانه ایجاد شود.

اغلب روش‌های نهان‌نگاری تطبیقی پیام محرمانه را بر اساس کمینه‌کردن تابع اعوجاج با هدف حفظ مشخصات پوشانه جاسازی می‌کنند $D: \mathcal{C} \times \mathcal{S} \rightarrow [0, \infty]$ که D بیانگر اعوجاج کل بوده و تقریبی از قابلیت تشخیص تغییرات ناشی از جاسازی پیام را منعکس می‌کند؛ بنابراین به کمینه‌رساندن آن باید باعث بهبود امنیت نهان‌نگاری می‌شود.

اعوجاج جاسازی به صورت $D(x, y)$ تعریف می‌شود که به هر دو تصویر پوشانه و حامل وابسته است. با فرض ثابت‌بودن پوشانه، می‌توان اعوجاج را به صورت $D(x, y) = D(y)$ ساده کرد. در اعوجاج جمع‌شونده،

تغییرات پیکسل‌های پوشانه مستقل فرض می‌شود؛ بنابراین هزینه تغییر هر پیکسل به تغییرات پیکسل‌های دیگر بستگی ندارد. پس اعوجاج کل از مجموع هزینه پیکسل‌های ویرایش شده به دست می‌آید:

$$D(y) = \sum_{i=1}^n \rho_i [x_i \neq y_i] \quad (2)$$

که $\rho_i = \rho(y_i) \in \mathbb{R}$ توابع غیر منفی و محدودی هستند که هزینه جایگزینی i امین پیکسل پوشانه x_i با y_i را تعیین می‌کنند. روند نهان‌سازی یک مسئله بهینه‌سازی است که هدف آن به کمینه‌رساندن اعوجاج کل با شرط ظرفیت جاسازی مشخص است؛ اما این مسئله به دلیل مشخص‌نبودن مکان تغییرات، از نظر محاسباتی غیرممکن است. از این‌رو، به جای اعوجاج کل، میانگین اعوجاج کمینه می‌شود؛ بنابراین در این روش $H(\pi)$ بیت از پیام با میانگین اعوجاج $E_\pi(D)$ با احتمال $\pi(y)$ در تصویر پوشانه جاسازی می‌شود:

$$E_\pi(D) = \sum_{i=1}^n \rho_i \pi_i \quad (3)$$

$$H(\pi) = - \sum_{i=1}^n \pi_i \log \pi_i \quad (4)$$

ظرفیت جاسازی مورد انتظار بیشینه آنتروپی بوده و $H_k(\pi_1, \pi_2, \dots, \pi_k)$ برای تابع آنتروپی به صورت $\sum_{i=1}^k \pi_i = 1$ تعریف می‌شود. به‌طور معمول ظرفیت جاسازی محدود فرض شده و هدف کمینه‌کردن میانگین اعوجاج برای جاسازی یک پیام به طول L است که می‌تواند به صورت یک مسئله بهینه‌سازی زیر فرموله شود:

$$\min_{\pi} E_\pi(D) \quad (5) \\ s.t: H(\pi) = L$$

درحالی‌که I_i اطلاعات فیشر نهان‌نگاری^۴ است :

$$I_i = \int_{\mathbb{R}} \frac{1}{p_i^{(c)}(x)} \left(\frac{\partial p_i^{(s)}(x)}{\partial \pi_i} \right) \bigg|_{\pi_i=0}^2 \quad (13)$$

با فرض ظرفیت جاسازی محدود، هدف کمینه‌کردن میانگین اعوجاج برای جاسازی یک پیام به طول L است که می‌تواند به‌صورت یک مسئله بهینه‌سازی فرموله شود:

$$\min \sum_{i=1}^n I_i \pi_i^2 \\ s.t.: H(\pi) = L \quad (14)$$

که باید برای هر پیکسل i به‌صورت عددی حل شود. حل این معادله برای π_i معادل یافتن x است که رابطه $\lambda I_i(0)/2 = x \ln(x-2)$ را ارضا کند، درحالی‌که $x = \pi_i^{-1} \geq 3$ زیرا $H(x)$ برای $x = 1/3$ به بیشینه می‌رسد. برای حل سریع این معادله به‌ازای همه پیکسل‌ها، تابع معکوس $y = x \ln(x-2)$ برای $y \leq 10^3$ جدول‌بندی‌شده و یک راه‌حل تکراری مجانبی برای $y > 10^3$ اجرا می‌شود. برای کمینه بودن π_i محاسبه‌شده، دومین مشتق تابع هدف نسبت به π_i باید مثبت باشد که به معنی $\lambda > 0$ است. ضریب لاگرانژ λ از شرط ظرفیت (۴) محاسبه و پس از محاسبه احتمالات تغییر بهینه π_i ، هزینه جاسازی هر پیکسل از رابطه (۱۰) محاسبه می‌شود.

۴-۲- اعوجاج غیر جمع‌شونده

تابع اعوجاج برای امنیت نهان‌نگاری بسیار حائز اهمیت است. بیشتر روش‌های مبتنی بر اعوجاج، احتمالات تغییر پیکسل‌های پوشانه را مستقل فرض می‌کنند و تعامل بین تغییرات جاسازی در نظر گرفته نمی‌شود؛ اما در واقعیت، تغییرات پیکسل‌های مجاور در تصاویر طبیعی بر یکدیگر مؤثر هستند؛ از این‌رو، لحاظ‌کردن این وابستگی در عمل منجر به کاهش بهینگی می‌شود. این واقعیت، ایده اصلی یک طرح اعوجاج غیرجمع‌شونده است که با لحاظ‌کردن اثر متقابل تغییرات محلی، از تولید الگوهای قابل‌شناسایی توسط نهان‌کاو جلوگیری کرده و محرمانگی را بهبود می‌بخشد.

تأثیر جاسازی را می‌توان با تابعی مدل کرد که فاصله بین تصویر پوشانه و تصویر حامل را در فضای ویژگی یا فضای توصیف دیگر اندازه‌گیری کند. پس تابع اعوجاج D را می‌توان به‌صورت زیر تعریف کرد:

$$D(x, y) = \|f(x) - f(y)\| \quad (15)$$

برای یک ظرفیت مشخص، کمینه اعوجاج می‌تواند با پیروی تغییرات جاسازی از توزیع گیبز^۱ به دست آید. طبق اصل آنتروپی بیشینه^۲، احتمال جاسازی بهینه π به فرم توزیع گیبز بیان می‌شود:

$$\pi_\lambda(y) = \frac{1}{Z(\lambda)} \exp(-\lambda D(y)) \quad (6)$$

که $Z(\lambda)$ فاکتور نرمال‌سازی است به‌طوری‌که:

$$Z(\lambda) = \sum_{y \in Y} \exp(-\lambda D(y)) \quad (7)$$

احتمال تغییر برای هر عنصر پوشانه به دست می‌آید؛ بنابراین، احتمال تغییرات $\pi_\lambda(y)$ می‌تواند به‌صورت ضرب احتمالات حاشیه‌ای تغییرات پیکسل‌های فردی فرموله شود:

$$\pi_\lambda(y) = \prod_{i=1}^n \pi(y_i) \quad (8)$$

$$\pi(y_i) = \frac{\exp(-\lambda \rho_i)}{\sum_{y_i \in I_i} \exp(-\lambda \rho_i)} \quad (9)$$

پارامتر عددی $\lambda > 0$ شرط آنتروپی را ارضا می‌کند (۴) و برای طول پیام مشخص L ، می‌تواند با جستجوی دودویی محاسبه شود. طبق تغییرات $\lambda \in (0, \infty)$ ، می‌توان یک رابطه بین حداکثر آنتروپی و حداقل اعوجاج حاصل از جاسازی فرض کرد که محدوده نرخ اعوجاج^۳ نامیده می‌شود و برای شبیه‌سازی یک طرح جاسازی بهینه مفید است.

برای به دست‌آوردن هزینه تغییرات پیکسل‌ها برای جاسازی عملی، می‌توان معادله (۹) را با انتخاب دلخواه λ معکوس کرد؛ بنابراین، برای جاسازی سه‌گانه، هزینه‌ها به‌صورت زیر محاسبه می‌شوند:

$$\rho_i = \ln(1/\pi_i - 2) \quad (10)$$

۴-۱-۲- رویکرد مبتنی بر مدل

در این رویکرد به‌جای تعریف هزینه‌های اکتشافی از یک مدل آماری استفاده می‌شود. با فرض $p_i^{(c)}(x)$ به‌عنوان توزیع احتمال مدل پوشانه در مؤلفه i ام و استفاده از الگوریتم جاسازی LSBM، می‌توان توزیع حامل را به‌صورت زیر مدل‌سازی کرد:

$$p_i^{(s)}(x) = (1 - 2\pi_i)p_i^{(c)}(x) + \pi_i p_i^{(c)}(x+1) + \pi_i p_i^{(c)}(x-1) \quad (11)$$

برای نرخ‌های تغییر کوچک، واگرایی KL به‌خوبی با عبارت درجه دوم تقریب زده می‌شود:

$$\sum_{i=1}^n D_{KL}(p_i^{(c)}, p_i^{(s)}) \approx \sum_{i=1}^n I_i \pi_i^2 \quad (12)$$

¹ Gibbs distribution

² Principle of maximum entropy

³ Rate-distortion bound

⁴ Steganographic fisher information

که $f(x)$ و $f(y)$ به ترتیب ویژگی‌های بردار پوشانه و حامل را توصیف می‌کنند که با کمک فیلترهای محلی و اغلب بر اساس رخداد مقادیر پیکسل محاسبه می‌شوند. از این‌رو، تابع اعوجاج بیان‌شده در این معادله غیرجمع‌شونده است؛ اما به دلیل پیچیدگی مسئله کمینه‌سازی اعوجاج غیرجمع‌شونده و عدم وجود روش‌های جاسازی عملی مناسب برای این روش‌ها، پژوهش‌گران به‌طور معمول از دو رویکرد تقریب به اعوجاج جمع‌شونده^۱ یا بزرگ‌نمایی با اعوجاج محدود^۲ استفاده می‌کنند.

در رویکرد تقریب، با فرض عدم تأثیر تغییر یک پیکسل بر قابلیت شناسایی پیکسل‌های همسایه، تابع اعوجاج غیرجمع‌شونده D با یک اعوجاج جمع‌شونده تقریب زده می‌شود به‌طوری‌که اثر تغییر ناشی از جاسازی برای هر پیکسل به‌صورت مجزا ارزیابی می‌شود $\hat{D}(y) = \sum_{i=1}^n D(y_i X_{\sim i})$. از آنجایی‌که در واقعیت، اعوجاج کل به دلیل وجود تعامل بین پیکسل‌ها با مجموع اعوجاج پیکسل‌ها برابر نیست پس تقریب جمع‌شونده باعث از دست رفتن ظرفیت، به‌ویژه برای ظرفیت‌های بالا می‌شود؛ بنابراین، تقریب جمع‌شونده نمی‌تواند فعل و انفعالات تغییرات مجاور را لحاظ کند و باعث کاهش محرمانگی می‌شود. برای حل این مشکل، در رویکرد دوم، مسئله اصلی به زیرمسائل ساده محلی که حل آن‌ها نسبتاً آسان‌تر است، تقسیم می‌شود.

در رویکرد اعوجاج محدود از دو فرض زیر برای لحاظ کردن همبستگی تغییرات پیکسل‌ها در تابع اعوجاج بهره گرفته می‌شود:

- اختصاص هم‌زمان هزینه‌ها ضرورتی ندارد.
- افزایش و کاهش مقدار پیکسل لزوماً هزینه یکسانی ندارند.

در [۲۵] نویسندگان یک چارچوب کلی با تابع اعوجاج غیرجمع‌شونده مبتنی بر ساختار گیبز پیشنهاد کرده‌اند. در این روش اعوجاج کل به‌صورت مجموع پتانسیل‌های محلی بر روی گروهک^۳ (گروه کوچکی از پیکسل‌ها) $D(y) = \sum_{c \in C} V_c(y)$ تعریف می‌شود که اعوجاج محدود نامیده می‌شود، درحالی‌که دو مؤلفه هر گروهک همسایه هستند. در این تابع اعوجاج با بررسی تغییرات پیکسل‌های مجاور، همگام‌شدن تغییرات

پیکسل‌های هر گروهک را تشویق می‌کند زیرا همگام‌سازی مکان و جهت ویرایش‌های ناشی از جاسازی، نواحی یک‌پارچه‌تری از تغییرات را ایجاد می‌کند؛ در این صورت فقط پیکسل‌های مرزی بر خصوصیات آماری تصویر به‌دست‌آمده تأثیر می‌گذارند که باعث بهبود محرمانگی و امنیت می‌شود. در این روش تصویر پوشانه با گروه‌بندی پیکسل‌ها به چند زیرتصویر تقسیم می‌شود؛ به‌طوری‌که در هر زیر تصویر پیکسل‌ها مستقل از یکدیگر هستند؛ سپس پیام به قسمت‌های کوچکی تقسیم می‌شود که هر قسمت باید در یک زیرتصویر جاسازی شود به‌طوری‌که اعوجاج را به کمینه برساند. در این صورت، می‌توان یک طرح کدگذاری عملی برای جاسازی پیام در هر زیر تصویر اعمال کرد. از آنجاکه ویژگی‌های نهان‌کاوی را می‌توان به‌عنوان مجموع ویژگی‌های محلی در تصویر نشان داد، یک حد بالا برای این ویژگی‌های محلی می‌تواند در اعوجاج محدود تعریف شود.

۴-۳- اولویت جاسازی

از آنجاکه نهان‌نگاری بر اساس ویرایش پوشانه، خصوصیات آماری پوشانه را در حین جاسازی پیام تغییر می‌دهد؛ بنابراین با توجه به تأثیر جاسازی بر محرمانگی، لازم است مناطق امن پوشانه انتخاب شوند تا در صورت تغییر، تأثیر قابل‌توجهی بر روی مشخصات کلی پوشانه نداشته باشند. در این راستا تاکنون قوانین مختلفی برای تعیین اولویت ویرایش و انتساب هزینه به پیکسل‌های تصویر ارائه شده است [۲۶] که در اینجا به اختصار برخی از قوانین مؤثر را بیان می‌کنیم:

۱. پیچیدگی^۴ یک اولویت اساسی برای انتساب هزینه^۵ در نهان‌نگاری است، بدین معنی که نواحی پیچیده تصویر باید اولویت بالاتری نسبت به مناطق هموار برای جاسازی داشته باشند؛ زیرا مدل‌سازی بافت‌های غیرمتناوب و نواحی نویزی دشوار بوده و تغییرات در این نواحی منجر به اعوجاج جزئی در فضای ویژگی نهان‌کاو و در نتیجه کاهش احتمال شناسایی آن خواهد شد. تخصیص هزینه در بیشتر روش‌های نهان‌نگاری از اولویت پیچیدگی پیروی می‌کند [27] HUGO، [28] S-WOW، [29] UNIWARD، [30] MVG و [31] MiPOD.

⁴ Complexity

⁵ Cost assignment

¹ Approximation to additive distortion

² Majorization by bounding distortion

³ Clique

۲. انتشار هزینه^۱: با توجه به این قانون، هزینه تغییر دو مؤلفه مجاور نباید اختلاف زیادی با هم داشته باشند. به بیان دیگر، هر مؤلفه با اولویت تغییر بالا (اولویت تغییر پایین) باید رتبه خود را به پیکسل‌های مجاور خود نشر دهد. با اعمال این قانون، هزینه‌ها در مناطق همسایگی محلی هموار می‌شوند؛ پس اگر دو پیکسل اولویت پیچیدگی یکسانی داشته باشند بر اساس این قانون پیکسل قرار گرفته در ناحیه با بافت پیچیده‌تر نسبت به پیکسل دیگر هزینه پایین‌تری به دست خواهد آورد. این قانون ممکن است آنتروپی^۲ جاسازی را در مناطق با بافت غنی افزایش و از سوی دیگر قابلیت تشخیص را کاهش دهد. این قانون برای نخستین بار در روش [33] HILL با موفقیت مورد استفاده قرار گرفت.

۳. خوشه‌بندی^۳: در این قانون ادعا می‌شود که بهتر است تغییرات جاسازی به صورت خوشه‌ای و در کنار هم انجام شوند. تغییرات خوشه‌ای نسبت به تغییرات پراکنده باعث اعوجاج کمتری در فضای ویژگی نهان‌کاو می‌شود، این بدان معنا است که خوشه‌بندی تغییرات جاسازی در پیکسل‌های مجاور، مقادیر باقیمانده^۴ کمتری را نسبت به زمانی که تغییرات در پیکسل‌های پراکنده رخ می‌دهند، به وجود می‌آورد؛ بنابراین اگر چند پیکسل پس از اعمال دو قانون قبل هنوز دارای هزینه مشابه باشند، اولویت تغییر با پیکسل‌هایی است که در یک خوشه قرار دارند.

۴. همگام‌سازی جهت تغییرات^۵: این قانون بیان می‌کند که تغییرات در جهت یکسان برای پیکسل‌های مجاور موجب اعوجاج کمتر پوشانه و در نتیجه قابلیت شناسایی کمتر حامل می‌شود؛ بنابراین اگر تغییر پیکسل‌های مجاور هماهنگ باشند، فقط پیکسل‌های مرزی بر تصویر باقیمانده استخراج شده تأثیر می‌گذارند که قابلیت تشخیص را کاهش می‌دهد. لی و همکاران [۳۴] و دنمارک و همکاران [۳۵] به طور مستقل از این قانون برای تعریف تابع اعوجاج بهره گرفتند؛ سپس ژانگ و همکاران [۳۶] چارچوبی جدید در این راستا ارائه

دادند. به تازگی در [۳۷-۳۹] راهکارهای مناسبی با الهام از میدان تصادفی مارکوف ارائه شده است. ۵. استفاده از اطلاعات جانبی^۶: در این قانون فرستنده از یک پوشانه اولیه^۷ که قبل از جاسازی پیام تحت نوعی پردازش قرار می‌گیرد، برای تعدیل هزینه‌های تغییرات جاسازی کمک می‌گیرد [۳۰-۴۳]. نمونه‌هایی از چنین پردازش‌هایی عبارتند از تغییر اندازه، فشرده‌سازی JPEG، تبدیل از رنگ به مقیاس خاکستری و غیره.

در دو قانون نخست احتمال مساوی برای تغییر مثبت و منفی در جاسازی سه‌گانه به دست می‌آید و جاسازی به صورت متقارن انجام می‌شود و به همین دلیل بیشینه آنتروپی حاصل می‌شود. با این وجود، در نظر گرفتن تعاملات بین تغییرات جاسازی در یک همسایگی محلی (قانون سوم و چهارم) و استفاده از اطلاعات جانبی (قانون پنجم) می‌توانند در اصلاح هزینه‌های جاسازی مفید واقع و باعث بهبود امنیت و محرمانگی شوند؛ اما جاسازی در این روش‌ها به صورت نامتقارن انجام می‌شود و احتمالات مثبت و منفی متفاوت برای هر پیکسل به دست می‌آید. این روش‌ها به دلیل پیروی از آنتروپی شرطی نمی‌توانند بیشینه پیام را جاسازی کنند؛ اما به دلیل ناچیز بودن این افت ظرفیت، می‌توان آن را در برابر قابلیت تشخیص کمتر و امنیت بیشتر نادیده گرفت. تعاملات بین تغییرات حاصل از جاسازی را می‌توان با استفاده از توابع اعوجاج غیر جمع‌شونده لحاظ کرد که اعوجاج به صورت مجموع پتانسیل‌های محلی تعریف می‌شود و خوشه‌بندی و همگام‌سازی جهت تغییرات مجاور را تشویق می‌کند.

۴-۴- جاسازی پیام

ساده‌ترین راه برای پنهان کردن پیام در یک تصویر در مقیاس خاکستری، با فرض اعمال تغییرات جزئی در تصویر، جایگزینی کم‌اهمیت‌ترین بیت^۸ یا به اختصار LSBR است که شامل جایگزینی بیت‌های کم‌اهمیت (LSB) از پیکسل‌های انتخاب‌شده بر اساس نقشه هزینه با بیت‌های پیام موردنظر می‌شود. این روش هر پیکسل را حداکثر یک بیت تغییر می‌دهد، تغییری که برای

¹ Cost spreading rule

² Entropy

³ Clustering rule

⁴ Residual

⁵ Synchronizing modification directions rule

⁶ side-information

⁷ precover

⁸ Least Significant Bit Replacement

چشم انسان قابل مشاهده نیست؛ زیرا سامانه بینایی انسان نمی‌تواند تغییرات کم را تشخیص دهد. باین‌حال، با توجه به سادگی روش، تغییر حتی قسمت کوچکی از پیکسل‌ها به‌طور قابل‌توجهی توزیع آماری آن را تغییر می‌دهد و شناسایی را ساده می‌کند. روش تطبیق کم‌اهمیت‌ترین بیت^۱ یا LSBM روش قبل را توسعه داده و مقدار پیکسل منتخب برای جاسازی پیام را بر اساس مقدار بیت پیام، بدون تغییر گذاشته و یا به‌طور تصادفی یک یا چند واحد افزایش یا کاهش می‌دهد. LSBR میانگین پیکسل‌ها را تغییر می‌دهد؛ درحالی‌که LSBM میانگین را حفظ می‌کند، اما واریانس را افزایش می‌دهد. LSBM به‌عنوان پایه الگوریتم‌های جاسازی در اغلب روش‌های نهان‌نگاری تصویر مورد استفاده قرار می‌گیرد. با توجه به اهمیت فرایند جاسازی در امنیت نهان‌نگاری، فیلر و همکاران [۵۲] یک الگوریتم کدگذاری عملی (STC^۲) را پیشنهاد دادند که در نزدیکی محدوده ظرفیت-اعوجاج^۳ جاسازی را انجام می‌دهد. در روش‌های مبتنی بر اعوجاج جمع‌شونده، الگوریتم STC می‌تواند برای جاسازی در نزدیکی محدوده اعوجاج نظری استفاده شود؛ از این‌رو، از امنیت به نسبت خوبی در برابر ابزارهای نهان‌کاوی برخوردار است؛ بنابراین پژوهش‌گران فرایند جاسازی پیام را به STC واگذار کرده و توجه خود را به بهبود امنیت تابع اعوجاج معطوف می‌سازند.

۴-۴-۱- جاسازی عملی

در واقع STC یک روش کدگذاری موازی با کدهای تصحیح خطای کانولوشنی است و الگوریتم رمزگشایی آن (الگوریتم viterbi) بر اساس الگوریتم داربست^۴ عمل می‌کند. به‌منظور یافتن بردار حامل با کمترین اعوجاج، STC از یک گراف استفاده می‌کند که شامل رؤس مربوطه به مقادیر احتمالی v_s است، درحالی‌که لبه‌ها هزینه‌های مشخص‌شده در نقشه هزینه را نشان می‌دهند. سندرم s به‌صورت $s = H \cdot v_s = m$ در هنگام دریافت پیام m تعریف می‌شود. برای این منظور، از یک ماتریس بررسی برابری^۵ H استفاده می‌شود که با پر کردن قطر یک ماتریس پراکنده با تکرار ماتریس H با ابعاد $w \times h$ به دست می‌آید. همان‌طور که در شکل ۴

^۱ Least Significant Bit Matching

^۲ Syndrome-trellis codes

^۳ Payload-distortion bound

^۴ Trellis

^۵ Check-parity matrix

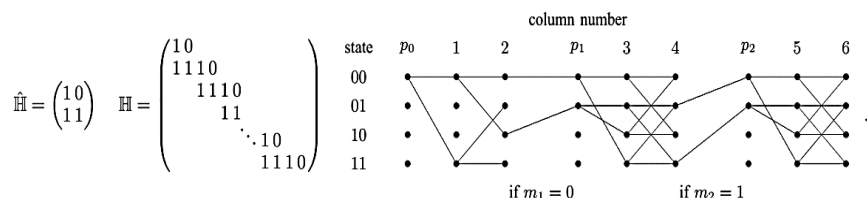
نشان داده شده است، n نسخه از زیر ماتریس‌ها یکی در کنار دیگری قرار گرفته و یک خط به پایین منتقل می‌شود. بقیه ماتریس با صفر پر شده است. سپس گراف با هدف یافتن بردار v_s به‌عنوان بردار حامل که $H \cdot v_s = m$ را برآورده کند، پیمایش می‌شود. هر بردار v_s به‌عنوان مسیری که از سمت چپ به راست گراف می‌رود، نشان داده می‌شود. مسیرها با توجه به هزینه‌های موجود در نقشه هزینه، ادامه می‌یابند و یا متوقف می‌شوند. راه‌حل بهینه بر اساس کمترین هزینه (کوتاه‌ترین مسیر) ارائه می‌شود. در حال حاضر، رویکرد داربست مؤثرترین روش عملی جاسازی از نظر کارایی بوده و نزدیک‌ترین روش به محدوده نظری است که از طریق شبیه‌ساز به دست می‌آید. به همین دلیل در اغلب روش‌های نهان‌نگاری مورد استفاده قرار می‌گیرند. STC به‌راحتی به کدگذاری غیر دودویی گسترش می‌یابد. جزئیات بیشتر در [۵۲] بیان شده است.

۴-۴-۲- شبیه‌سازی جاسازی بهینه

استفاده از یک روش کدگذاری مانند STC منجر به زیربهینگی^۶ جزئی در جاسازی می‌شود که می‌توان با شبیه‌سازی جاسازی بهینه این امر را مدیریت کرد. کمینه اعوجاج مورد انتظار برای یک ظرفیت ثابت، از رابطه $\min \sum_{i=0}^n \rho_i \pi_i$ به دست می‌آید، درحالی‌که π_i بیان‌گر احتمال تغییر هر پیکسل i است که به‌طورمستقیم با مقدار ρ_i مرتبط است و از رابطه (۹) محاسبه می‌شود.

با شبیه‌سازی جاسازی بهینه، پژوهش‌گران می‌توانند توابع اعوجاج پیشنهادی خود را بدون اتکا به الگوریتم‌های عملی کدگذاری طراحی و آزمایش کنند. علاوه بر این، می‌توان از شبیه‌ساز به‌عنوان حد بالا برای مقایسه کارایی طرح‌های کدگذاری استفاده نمود. در این روش برای جلوگیری از هرگونه سوگیری در طول شبیه‌سازی فرایند جاسازی، به‌جای پیام واقعی، پیام محرمانه دنباله‌ای از بیت‌های تصادفی غیرهمبسته فرض می‌شوند؛ بنابراین، تنها دانش مورد نیاز در مورد پیام برای بررسی تابع اعوجاج، طول پیام است. پس از نهایی شدن تابع اعوجاج، برای جاسازی پیام اصلی می‌توان از روش کدگذاری عملی مانند STC استفاده کرد.

^۶ Sub-optimality



شکل ۴: روش کدگذاری STC [52]

Figure 4: STC coding method [52]

ویژگی استخراج‌شده از تصاویر پوشانه و حامل در یک فضای ویژگی با ابعاد بالا محاسبه می‌شود. در این روش هزینه پیکسل با تغییر دامنه ویژگی‌های ماتریس همسایگی پیکسل تفریق‌کننده^۵ [53] (SPAM) به دست می‌آید، از این‌رو، پیکسل با هزینه کمتر باعث افزایش اعوجاج در بردارهای ویژگی می‌شود و بیشترین تغییرات جاسازی‌شده در مناطق با بافت پیچیده رخ می‌دهد. اگر پیکسل در هر جهتی بتواند به‌طور دقیق مدل‌سازی شود، باید به‌عنوان یک نقطه هموار در نظر گرفته شود و هزینه بیشتری را به آن اختصاص دهد. ویژگی‌های SPAM از همبستگی بین بقایای پیش‌بینی‌شده پیکسل‌های همسایه به دست می‌آیند. HUGO با یک نهان‌کاو با ویژگی‌هایی با ابعاد بالاتر مانند مدل غنی مکانی^۶ [54] (SRM) قابل تشخیص است. ویژگی‌های SRM از بقایای پیش‌بینی‌شده در جهات مختلف محاسبه می‌شوند؛ بنابراین همبستگی بین پیکسل‌ها می‌تواند بیشتر مورد استفاده قرار گیرد. با این بینش و در راستای بهبود امنیت روش HUGO در مقابل ویژگی‌های SRM، روش وزن‌های موجک^۷ [28] (WOW) ارائه گردید که در این روش پیکسل‌هایی که توسط یک بانک از فیلترهای جهت‌دار قابل پیش‌بینی هستند، هزینه بالاتری به دست می‌آورند. سپس، الگوریتم اعوجاج مرجع موجک جهانی [29] (S-UNIWARD)^۸ برای توسعه تابع هزینه WOW پیشنهاد شد تا آن را ساده‌تر کرده و برای جاسازی در دامنه مناسب تعمیم دهد. در سال ۲۰۱۴، لی و همکاران روش [33] HILL^۹ را برای بهبود عملکرد روش‌های قبلی معرفی کردند که با انتشار هزینه‌ها با استفاده از یک فیلتر پایین‌گذر باعث می‌شود تغییرات جاسازی‌شده در نواحی پیچیده‌تر تصویر اعمال شود.

۵- مرور روش‌های نهان‌نگاری سنتی

۵-۱- مرور روش‌های مبتنی بر اعوجاج

جمع‌شونده

روش‌های نهان‌نگاری از دو اصل مختلف برای طراحی پیروی می‌کنند: حفظ مدل و به کمینه‌رساندن اعوجاج. روش‌های مبتنی بر مدل، یک مدل آماری به پوشانه اختصاص می‌دهند و سعی می‌کنند مدل را با به کمینه‌رساندن واگرایی KL بین تصاویر پوشانه و حامل، حفظ کنند. فریدریش و همکاران نخستین رویکرد مدل محور را ارائه دادند که در آن از توزیع چندمتغیره گاوسی^۱ (MVG) [30] یا توزیع چند متغیره تعمیم‌یافته گاوسی^۲ (MVGG) [32] برای مدل‌سازی باقیمانده نوفه پیکسل‌ها، به‌عنوان دنباله‌ای از متغیرهای تصادفی گاوسی چندمتغیره با واریانس‌های مختلف، استفاده شده است. صدیقی و همکاران [۳۱] یک رویکرد پیشرفته مبتنی بر مدل را بر اساس به کمینه‌رساندن قدرت ردیاب بهینه^۳ (MiPOD) ارائه دادند که به امنیت بهتری در برابر سایر روش‌های نهان‌نگاری مبتنی بر مدل دست یافته است؛ اما از آنجاکه اتخاذ یک مدل کارآمد دشوار است و عدم تطابق مدل باعث افزایش قابلیت تشخیص و کاهش امنیت می‌شود، اغلب روش‌های نهان‌نگاری به سمت کمینه‌سازی تابع اعوجاج هدایت می‌شوند.

از آنجاکه طراحی توابع اعوجاج به اصول اکتشافی متکی است، هیچ ارتباط مستقیمی بین اعوجاج و قابلیت تشخیص آماری وجود ندارد؛ اما فرستنده می‌تواند این ارتباط را با تعریف تابع اعوجاج در قالب خطای تصمیم‌گیری نهان‌کاو ایجاد کند.

نخستین روش مبتنی بر اعوجاج، روش حامل قویاً غیرقابل تشخیص [27] HUGO^۴ است که در این روش اعوجاج به‌عنوان مجموع وزنی اختلافات بین بردارهای

^۵ Subtractive Pixel Adjacency Model

^۶ Spatial Rich Model

^۷ Wavelet Obtained Weights

^۸ Spatial UNiversal WAVElet Relative Distortion

^۹ High Low Low

^۱ Multivariate Gaussian

^۲ Multivariate Generalized Gaussian

^۳ Minimizing the Power of Optimal Detector

^۴ Highly Undetectable steGO

الگوریتم HUGO در سال ۲۰۱۰ پیشنهاد شد [۲۷] و در همان سال به عنوان الگوریتم جاسازی شده برای BOSS استفاده شد که به منظور به کمینه رساندن اعوجاج در یک فضای مشخصه با ابعاد بالا که از اختلاف چهار پیکسل همسایه به دست آمده، طراحی گردید. نقشه هزینه الگوریتم HUGO به گونه ای تعریف شده است که هزینه ویرایش پیکسل از رابطه زیر به دست می آید:

$$D(x, y) = \|f(x) - f(y)\| = \sum_{j=1}^d w_j |f_j(x) - f_j(y)| \quad (۱۶)$$

که f یک بردار ویژگی است که از ماتریس هم زمانی^۱ (ویژگی های SPAM) تولید می شود و $f(x)$ به بردار ویژگی d بعدی تصویر x اشاره دارد. n بیانگر تعداد پیکسل ها و d تعداد ویژگی ها است؛ از این رو، هر سطر از این ماتریس شامل تعداد وقایع سه گانه مقادیر (d_1, d_2, d_3) در تصویر باقیمانده است که با فیلتر کردن تصویر با هسته $[-1, 1]$ و به دنبال آن کوتاه سازی^۲ در فاصله $[-T, \dots, T]$ به دست می آید. w_j وزنی است که با سه گانه $(d_1, d_2, d_3) \in [-T, \dots, T]$ مرتبط است. از این رو:

$$D(x, y) = \sum_{d_1, d_2, d_3 = -T}^T w_{d_1, d_2, d_3} |f_{d_1, d_2, d_3}(x) - f_{d_1, d_2, d_3}(y)|$$

که f_{d_1, d_2, d_3} تفاوت (d_1, d_2, d_3) بین پیکسل های همسایگی است. تخمین جمع شونده به صورت زیر خواهد بود:

$$D'(x, y) = \sum_{i=1}^n D(x, y_i | x_{\sim i}) |x_i - y_i| \quad (۱۸)$$

۵-۱-۲- الگوریتم موجک (WOW)

الگوریتم WOW یک طرح منطبق با محتوا با استفاده از بانک های فیلتر موجک برای ارزیابی اعوجاج جاسازی شده ارائه می دهد [۲۸] که برخلاف HUGO، به طور خاص برای جلوگیری از ایجاد تغییرات جاسازی شده در محتوای قابل مدل سازی طراحی شده است. الگوریتم WOW با استفاده از یک بانک فیلترهای بالاگذر جهت دار، باقیمانده جهت دار را به دست می آورد که محتوای اطراف هر پیکسل را در چندین جهت مختلف ارزیابی می کند. WOW با اندازه گیری تأثیر جاسازی بر روی هر باقیمانده جهت دار، اعوجاج را به گونه ای تعریف

می کند که محتوای آن حداقل در یک جهت (مناطق صاف و لبه های تمیز) قابل پیش بینی باشد. الگوریتم حاصل در برابر نهان کاوهای که از مدل های غنی [۳۲] برای استخراج ویژگی استفاده می کنند، نسبت به HUGO مقاوم تر است.

پس به اختصار:

۱- جاسازی در مناطقی با لبه در جهت افقی، عمودی و مورب انجام می شود.

۲- سه فیلتر جهت دار با هسته 8×8 ، $K^{(k)}$ ، $k \in \{h, v, d\}$ برای استخراج سه باقیمانده جهت دار $R^{(k)} = K^{(k)} * X$ استفاده می شود.

۳- مناسب بودن جاسازی با استفاده از رابطه $\xi^{(k)} = |R^{(k)}| * |K^{(k)}|$ به دست می آیند.

۴- هزینه جاسازی با استفاده از نرم reciprocal Hölder $\rho_{ij}^{(k)} = \left(\sum_{k=1}^3 |\xi_{ij}^{(k)}|^p\right)^{-p}$ به دست می آید. در حالی که $p = -1$.

۵-۱-۳- الگوریتم اعوجاج نسبی موجک در

حوزه مکان (S-UNIWARD)

S-UNIWARD یک الگوریتم جاسازی در حوزه مکان است که تابع اعوجاج آن در حوزه موجک تعریف شده است [۲۹]. تابع اعوجاج ارائه شده در این روش مشابه WOW اما برای جاسازی در یک دامنه دلخواه ساده تر و مناسب تر است. این اعوجاج از مجموع تغییرات نسبی ضرایب بین تصاویر حامل و پوشانه در حوزه موجک به دست می آید. جهت دار بودن باعث ایجاد تغییرات جاسازی در قسمت هایی از پوشانه مانند بافت ها یا مناطق نوفه ای می شود که مدل سازی آن ها در چند جهت دشوار است و از جاسازی در نواحی صاف یا لبه های تمیز جلوگیری می کند.

پس به اختصار:

۱- هزینه های جاسازی پیکسل از یک تابع اعوجاج به عنوان مجموع اختلاف مطلق نسبی بین ضرایب موجک تصاویر پوشانه و حامل به دست می آید.

۲- با فرض u, v امین ضریب موجک X در $k \in \{h, v, d\}$ زیر باند $W_{uv}^{(k)}(X)$ به صورت $W^{(k)} = K^{(k)} * X$ تعریف می شود.

۳- S-UNIWARD از همان هسته های تشکیل شده از موجک های 8-tap Daubechies مانند WOW استفاده می کند.

¹ Co-occurrence matrix

² Truncation

۴- اعوجاج غیرجمع‌شونده بین پوشانه X و حامل Y به صورت زیر تعریف می‌شود:

$$D(X, Y) = \sum_{k \in \{h, v, d\}} \sum_{u, v} \frac{|W_{uv}^{(k)}(X) - W_{uv}^{(k)}(Y)|}{\sigma + |W_{uv}^{(k)}(X)|} \quad (19)$$

$\sigma = 1$ ثابت تثبیت‌کننده است.

۵-۱-۴- الگوریتم فیلتر بالاگذر-پایین‌گذر-پایین‌گذر (HILL)

HILL یکی دیگر از الگوریتم‌های جاسازی در حوزه مکان است که در سال ۲۰۱۴ به عنوان نسخه ارتقا یافته WOW پیشنهاد شد [۳۳]. گاهی در مناطق بافت برخی پیکسل‌ها با هزینه بالا دیده می‌شود که به دلیل قابل پیش‌بینی بودن آن‌ها در یکی از جهات است. باین‌حال، برای یک پیکسل در یک منطقه بافت، حتی اگر در یکی از جهات قابل پیش‌بینی باشد، باید مقدار هزینه کمتری نسبت به مناطق هموار اختصاص داده شود. برای اطمینان از تخصیص هزینه پایین به پیکسل‌های مناطق بافت، در روش HILL سه هسته جهت‌دار با یک فیلتر بالا گذر غیر جهت‌دار که توسط دو فیلتر پایین‌گذر دنبال می‌شود، جایگزین می‌شود. از فیلتر بالاگذر برای مکان‌یابی نواحی با قابلیت تشخیص کمتر در تصویر استفاده می‌شود، درحالی‌که از دو فیلتر پایین‌گذر برای خوشه‌بندی مقادیر کم هزینه استفاده می‌شود.

پس به اختصار:

- ۱- از الگوریتم WOW سرچشمه گرفته است.
- ۲- سه هسته جهت‌دار با یک هسته H غیر جهت‌دار بالا گذر 3×3 جایگزین می‌شوند: $R = X * H$
- ۳- سپس هزینه پیکسل با استفاده از فرمول زیر محاسبه می‌شود:

$$(Y_0)\rho = \frac{1}{|R| * L_1} * L_2$$

- ۴- L_1 یک فیلتر میانگین 3×3 و L_2 یک فیلتر میانگین دیگر با ابعاد 15×15 است.

۵-۱-۵- الگوریتم توزیع چندمتغیره گوسی (MVG)

نخستین تلاش برای طراحی اعوجاج به عنوان یک کمیت مربوط به قابلیت تشخیص آماری در سال ۲۰۱۳ [۳۰] صورت گرفت. نویسندگان پیشنهاد کردند که از واگرایی KL بین توزیع‌های آماری تصاویر پوشانه و حامل، هنگام مدل‌سازی پیکسل‌های پوشانه به عنوان دنباله‌ای از متغیرهای تصادفی گاوسی مستقل با واریانس‌های نابرابر

گاوسی چند متغیره) استفاده شود. در این روش، با تحمیل یک مدل بر روی پوشانه و تخمین پارامترهای مدل، ابتدا واریانس محلی برای هر پیکسل محاسبه می‌شود؛ سپس، هزینه‌های ویرایش هر پیکسل به صورت تحلیلی از مدل برآورد شده محاسبه می‌شود به طوری‌که واگرایی KL بین توزیع پوشانه و حامل کمینه برسد. با فرض اینکه پیکسل‌ها دنباله‌ای از متغیرهای تصادفی مستقل اما نه لزوماً توزیع شده یکسان را تشکیل می‌دهند، نرخ تغییرات با استفاده از روش ضرب‌کننده‌های لاگرانژ قابل محاسبه است.

۱- پوشانه به عنوان دنباله‌ای از n متغیر تصادفی مستقل مدل می‌شود $X = (X_1, \dots, X_n)$ که با توزیع گاوسی با میانگین صفر کوانتیزه می‌شوند $p^{(i)} = (p_j^{(i)}), j \in \mathcal{M} \mid Q_{\Delta}(N(0, v_i))$

۲- فرایند جاسازی پیام، هر پیکسل را به طور مستقل با احتمال π_i ویرایش می‌سازد و پوشانه را به دنباله‌ای از متغیرهای تصادفی مستقل $Y = (Y_1, \dots, Y_n)$ به عنوان حامل با توزیع $q^{(i)}(\pi_i)$ $q_j^{(i)}, j \in \mathcal{M}$ تغییر می‌دهد. با افزایش π_i ، واگرایی KL بین پوشانه و حامل افزایش می‌یابد. برای نرخ‌های تغییر کوچک، واگرایی KL به خوبی عبارت درجه دوم اصلی آن تقریب زده می‌شود:

$$\sum_{i=1}^n D_{KL}(p^{(i)} \parallel q^{(i)}) \approx \sum_{i=1}^n \frac{1}{2} \pi_i^2 I_i(0) \quad (21)$$

که $I_i(0)$ اطلاع فیشر نهان‌نگاری است:

$$I_i(0) = \sum_j \frac{1}{p_j^{(i)}} \left(\frac{dq_j^{(i)}(\pi_i)}{d\pi_i} \Big|_{\pi_i=0} \right)^2 \quad (22)$$

۳- π_n با به کمینه‌رساندن واگرایی KL با محدودیت ظرفیت با استفاده از روش ضریب لاگرانژ تعیین می‌شود. در [۳۲]، رویکرد MVG با جایگزینی مدل با گاوسی تعمیم‌یافته، استفاده از یک الگوریتم محاسبه واریانس بهتر و بهره‌گیری از عملیات جاسازی پنج‌گانه گسترش یافت (MVGG).

۵-۱-۶- الگوریتم منطبق با محتوا با هدف به کمینه‌رساندن قابلیت تشخیص آماری (MiPOD)

MiPOD، در سال ۲۰۱۶، به عنوان یکی از امن‌ترین الگوریتم‌های جاسازی در حوزه مکان پیشنهاد شد [۳۱]. MiPOD یک روش نهان‌نگاری مبتنی بر مدل است که قدرت قوی‌ترین آشکارساز را که یک مهاجم می‌تواند در

هنگام مدل سازی باقی مانده نویز در یک تصویر به عنوان تعریف مستقل از متغیرهای تصادفی گاوسی با میانگین صفر و واریانس σ_i^2 برای هر مؤلفه پوشانه بسازد، به حداقل می رساند.

پس به اختصار:

$$\begin{aligned} 1- & r_n \sim \mathcal{N}(0, \sigma_n^2) = (p_{\sigma_n}(k))_{k \in \mathbb{Z}}^{-1} \\ p_{\sigma_n}(k) = \mathbb{P}(r_n = k) & \propto (2\pi\sigma_n^2)^{-1/2} \exp(-k^2/(2\sigma_n^2)) \end{aligned} \quad (23)$$

۲- مهاجم با یک آزمون فرضیه دودویی ساده مواجه است:

$$\begin{aligned} \mathcal{H}_0: & x_n \sim \mathcal{P}_{\sigma_n} \Rightarrow \varrho = \frac{\sqrt{2} \sum_{n=1}^N \pi_n \gamma_n \sigma_n^{-4}}{\sqrt{\sum_{n=1}^N \gamma_n^2 \sigma_n^{-4}}} \\ \mathcal{H}_1: & x_n \sim \mathcal{Q}_{\sigma_n, \gamma_n} \end{aligned} \quad (24)$$

۳- π_n با به کمینه رساندن اعوجاج ϱ با محدودیت ظرفیت با استفاده از روش ضریب لاگرانژ تعیین می شود.

۴- واریانس با استفاده از اتصالات محلی با فیلتر DCT دوبعدی با درجه ۸ در یک پنجره کشویی 9×9 تخمین زده می شود.

۵- اطلاعات فشر با فیلتر میانگین 7×7 فیلتر شده است.

۲-۵- مرور روش های مبتنی بر اعوجاج غیرجمع شونده

فیلر و همکاران [۲۵] نخستین تلاش را برای ارائه یک چارچوب غیرجمع شونده ارائه دادند و سپس روش اعوجاج محدودکننده HUGO¹ (HUGO-BD) را به عنوان یک اجرای عملی از این چارچوب معرفی کردند که از حد بالای تفاوت بین بردارهای ویژگی پوشانه و حامل به عنوان تابع اعوجاج استفاده می کند؛ سپس در سال ۲۰۱۵، همگام سازی جهت تغییرات (SMD) به عنوان پیشرفت در زمینه نهان نگاری معرفی شد. لی و همکاران [۳۴] و همچنین دنمارک و همکاران [۳۵] به طور مستقل رویکردهایی کارآمد برای تعیین اعوجاج غیرجمع شونده با به روزرسانی هزینه های جاسازی بر اساس تغییرات پیکسل های مجاور پیشنهاد کردند. آن ها تصویر پوشانه را به زیرتصاویر غیرهم پوشان تقسیم می کنند. از آنجاکه پیکسل های هر زیرتصویر مستقل از یکدیگر هستند، می توان از تخصیص هزینه و جاسازی با روش جمع شونده در هر زیرتصویر استفاده کرد. این

تقسیم بندی در [۳۵] با هدف اطمینان از استقلال تغییرات جاسازی در هر زیرتصویر انجام می شود و تغییرات مجاور ناهمگام را جریمه می کند، درحالی که در [۳۴] هدف خوشه بندی جهت تغییرات است و تغییرات جاسازی به صورت محلی به سمت جهت غالب در همسایگی جهت گیری می شوند. ژانگ و همکاران [۳۶] یک چارچوب غیرجمع شونده متفاوت ارائه داده اند که با تجزیه احتمال جاسازی توأم به دو احتمال جاسازی حاشیه ای و مشروط، عملیات جاسازی پیام را در دو مرحله به انجام می رساند. برای این منظور، تصویر پوشانه به بلوک های غیرهم پوشان تجزیه شده و اعوجاج توأم در داخل بلوک ها بر اساس قاعده زنجیره، به اعوجاج بر روی پیکسل های جداگانه تجزیه می شود. سو و همکاران [۳۷] یک چارچوب نهان نگاری مبتنی بر مدل معرفی می کنند که از میدان تصادفی گاوسی مارکوف^۲ (GMRF) برای تعریف تغییرات جاسازی متقارن در قالب یک مدل غیرجمع شونده بهره می گیرند. در [۳۹] یک چارچوب نهان نگاری غیر جمع شونده معرفی می شود که با استفاده از میدان تصادفی مارکوف همبستگی و تعاملات بین تغییرات پیکسل ها را مدل می کند. سپس، از استنتاج میدان متوسط جهت یافتن بهترین تقریب برای این مدل بهره می گیرد که تعاملات همسایگی را به عنوان یک اثر میانگین برآورده می کند.

۲-۵-۱- روش اعوجاج محدودکننده HUGO

در [۲۵] یک چارچوب کلی مبتنی بر تابع اعوجاج غیرجمع شونده معرفی شده است. در این چارچوب، ابتدا تصویر پوشانه به چندین زیرتصویر تجزیه می شود، به صورتی که پیکسل های داخل هر زیرتصویر با فاصله بیشتر از عرض پشتیبانی تابع پتانسیل از هم جدا می شوند لذا از هم مستقل خواهند بود. بر این اساس پیام نیز به چندین بخش تقسیم می شود. تخصیص هزینه و جاسازی پیام در زیرتصاویر به ترتیب و در هر زیرتصویر با روش جمع شونده و با توجه به زیرتصاویر جاسازی شده قبلی انجام می شود. اولین بخش از پیام در زیرتصویر اول با استفاده از یک روش جمع شونده مخفی می شود. سپس، هزینه پیکسل های زیرتصاویر بعدی با توجه به پیکسل های ویرایش شده، به روز می شوند. بخش های باقیمانده پیام به همین ترتیب در زیرتصاویر با هزینه های به روز شده جاسازی می شوند. به این ترتیب، می توان از روش های نهان نگاری جمع شونده موجود در

² Gaussian Markov Random Field

¹ HUGO Bounding Distortion

روی قابلیت تشخیص ندارد، زیرا هنگامی که یک مجموعه از پیکسل‌ها یا یک مقدار یکسان تغییر کنند، تنها پیکسل‌های مرزی به‌طور مستقیم در اعوجاج شرکت خواهند داشت. تخمین جمع‌شونده تابع اعوجاج (۲۵) به‌صورت زیر بیان می‌شود:

$$D_A(x, y) = \sum_{x_{ij} \neq y_{ij}} D(x_{ij}, y_{ij}) \quad (27)$$

که $D(x_{ij}, y_{ij})$ اعوجاج بین x و y_{ij} است. y_{ij} بیان‌گر تصویر x است که فقط i امین پیکسل x_{ij} به y_{ij} تغییر کرده است. در مقایسه با روش جمع‌شونده که هزینه پیکسل‌هایی که بدون تغییر باقی‌مانده‌اند صفر لحاظ می‌شود، در روش غیرجمع‌شونده، ممکن است به‌دلیل تأثیر پیکسل‌های مجاور هزینه غیر صفر برای آن‌ها به دست آید. هزینه مثبت عدم‌تغییر، علاوه‌بر افزایش ظرفیت جاسازی در یک پیکسل مشخص (آن‌تروپی)، تعداد تغییرات جاسازی را نیز افزایش می‌دهد. میزان اهمیت ظرفیت افزوده‌شده نسبت به افزایش نرخ تغییرات بستگی به این دارد که اعوجاج غیرجمع‌شونده تا چه میزان قابلیت تشخیص آماری را بهبود بخشد.

۵-۲-۳- روش خوشه‌بندی جهت تغییرات (CMD²)

در این روش [۳۴] تصویر پوشانه و پیام به $L_1 \times L_2$ زیربخش تقسیم می‌شوند؛ سپس یک ترتیب برای جاسازی در این زیر بخش‌ها انتخاب می‌شود. اعوجاج از تفاوت تصویر پوشانه و حامل به دست می‌آید $D = Y - X = (d_{ij})^{n_1 \times m_2}$. در ابتدا حامل برابر با پوشانه $Y = X$ در نظر گرفته می‌شود. هزینه‌های اولیه پیکسل‌های تصویر پوشانه $C = (c_{ij})^{n \times n_2}$ از طریق یکی از روش‌های جمع‌شونده متداول (مانند WOW، S-UNIWARD، HILL) به دست می‌آید. هزینه هر پیکسل در مکان (i, j) یک سه تایی $\rho_{ij} = (\rho_{ij}^+, \rho_{ij}^0, \rho_{ij}^-)$ است که ρ_{ij}^+ هزینه افزودن یک واحد به پیکسل، ρ_{ij}^- هزینه کاهش یک واحد از پیکسل و ρ_{ij}^0 هزینه عدم تغییر پیکسل است. برای زیربخش اول $\rho_{ij}^+ = \rho_{ij}^- = c_{ij}$ در نظر گرفته می‌شود. برای زیر بخش‌های دیگر هزینه بر اساس اعوجاج D ویرایش می‌شوند.

$$\rho_{ij}^+ = \begin{cases} c_{ij}/\alpha, & \text{if } \sum_{(l,j) \in N_{ij}} \delta(d_{l'j'} - 1) > \sum_{(l,j) \in N_{ij}} \delta(d_{l'j'}) \\ c_{ij}, & \text{otherwise} \end{cases} \quad (28)$$

جاسازی داده‌ها برای هر زیرتصویر استفاده کرد و تعاملات بین تغییرات جاسازی را نیز در نظر گرفت. در این روش، به دلیل تخمین هزینه‌های متفاوت برای تغییرات مثبت و منفی، تعادل بین احتمال افزایش و کاهش مقدار پیکسل از بین می‌رود.

با تکرار چندین بار رفت و برگشت جاسازی با کمک الگوریتم نمونه‌برداری گیبز، امید می‌رود که الگوی جاسازی‌شده به یک نمونه از جاسازی بهینه تبدیل شود. یک اجرای عملی برای این چارچوب به‌عنوان HUGO-BD (اعوجاج محدودکننده HUGO) معرفی شده است. این روش یک ویرایش از الگوریتم HUGO است که در آن اعوجاج غیرجمع‌شونده با استفاده از همسایگی محلی محاسبه می‌شود. در این روش HUGO با استفاده از ساختار گیبز با اعوجاج محدودکننده برای جاسازی سه‌گانه پیاده‌سازی می‌شود. معیار اعوجاج در الگوریتم HUGO-BD به‌عنوان یک نرم وزن‌دار در فضای ویژگی SPAM پیاده‌سازی می‌شود. عملکرد روش HUGO-BD به‌خصوص برای پیام‌های بزرگ بهتر از همتای جمع‌شونده خود، HUGO است اما نمی‌تواند از روش‌های دیگر پیشی بگیرد [۲۸، ۲۹، ۳۳].

۵-۲-۲- روش همگام‌سازی کانال انتخاب^۱ (Synch)

فرض کنید A یک روش نهان‌نگاری جمع‌شونده است که هر پیکسل را با حداکثر مقدار ± 1 تغییر می‌دهد. در این روش [۳۵]، هزینه ویرایش هر پیکسل پوشانه x_{ij} به $y_{ij} = x_{ij} + 1$ یا $y_{ij} = x_{ij} - 1$ یکسان و برابر با ρ_{ij} است؛ بنابراین، برای هر پیکسل $x_{ij} - y_{ij} \in \{-1, 0, +1\}$ با در نظر گرفتن C به‌عنوان مجموعه گروهک‌های دو پیکسلی که شامل پیکسل‌های دوتایی مجاور عمودی و افقی است، تابع اعوجاج غیرجمع‌شونده به شکل زیر تعریف می‌شود:

$$D(x, y) = \sum_{((ij),(kl)) \in C} S_c(x_{ij} - y_{ij}, x_{kl} - y_{kl}) \quad (25)$$

$$S_c = \begin{vmatrix} & 1 & 0 & -1 \\ 1 & 0 & A_c & vA_c \\ 0 & A_c & 0 & A_c \\ -1 & vA_c & A_c & 0 \end{vmatrix} \quad (26)$$

S_c یک ماتریس 3×3 است که به میانگین هزینه گروهک $A_c = (\rho_{ij} + \rho_{kl})/2$ وابسته است و $v \geq 0$ پارامتر کنترل‌کننده میزان زیان تغییرات ناهمگام است. مقدار صفر در قطر ماتریس S_c بدین معنی است که تغییر همگام پیکسل‌ها به $+1$ (یا -1) هیچ‌گونه تأثیری بر

² Clustering Modification Directions

¹ Synchronizing the selection channel

$$\rho_{ij} = \begin{cases} c_{ij}/\alpha, & \text{if } \sum_{(l,j) \in N_{ij}} \delta(d_{l'j'} - 1) < \sum_{(l',j) \in N_{ij}} \delta(d_{l'j'} + 1) \\ c_{ij}, & \text{otherwise} \end{cases} \quad (29)$$

که α عامل مقیاس‌گذاری، N_{ij} همسایگی چهارگانه پیکسل (i, j) و δ_z یک تابع نشان‌گر است که به صورت زیر تعریف می‌شود:

$$\delta(z) = \begin{cases} 1, & z = 0 \\ 0, & z \neq 0 \end{cases} \quad (30)$$

پس از تعیین هزینه هر پیکسل، جاسازی در هر یک از زیربخش‌ها با کمک شبیه‌ساز جاسازی بهینه و یا یک روش جاسازی عملی مانند STC انجام می‌شود.

۵-۲-۴- روش تجزیه اعوجاج توأم (DeJoin)

در این روش [۳۶]، اعوجاج به جای پیکسل بر روی بلاک پیکسل‌ها تعریف می‌شود. برای این منظور ابتدا تصویر پوشانه با n پیکسل به N بلاک مستقل تقسیم می‌شود. برای هر بلاک یک تابع اعوجاج توأم تعریف می‌شود و با هدف به کمینه‌رساندن مجموع اعوجاج بلاک‌ها، پیام در پوشانه جاسازی می‌شود. در این مدل، اعوجاج بین بلاک‌ها هنوز جمع‌شونده است؛ اما به دلیل اینکه تعداد الگوهای تغییر در هر بلاک زیاد است و موجب پیچیدگی محاسباتی زیادی می‌شود، نمی‌توان به طور مستقیم از STC استفاده کرد. برای کاهش پیچیدگی فرایند جاسازی، اعوجاج مشترک هر بلاک به اعوجاج روی پیکسل‌های مجزا تجزیه می‌شود و سپس STC می‌تواند به طور مناسبی جهت جاسازی پیام مورد استفاده قرار گیرد.

در این روش ابعاد هر بلاک 2×1 در نظر گرفته می‌شود. با توجه به اینکه هر پیکسل سه الگوی تغییر $(+1, 0, -1)$ دارد، هر بلاک شامل ۹ الگوی تغییر خواهد بود. برای کاهش پیچیدگی فرایند جاسازی، یک روش جاسازی دومرحله‌ای پیشنهاد می‌شود که احتمال مشترک π^i به دو احتمال حاشیه‌ای و شرطی تجزیه می‌شود. ابتدا احتمال حاشیه‌ای بر پیکسل اول هر بلاک محاسبه می‌شود، سپس احتمال شرطی برای بیان احتمال تغییر پیکسل دوم به شرط یافتن پیکسل اول تعریف می‌شود. برای طول L شخص L ، در مرحله اول $L_1 = \sum_{i=1}^N H_3(\pi_1^i)$ از پیام در پیکسل‌های اول هر بلاک و در مرحله دوم $L_2 = \sum_{i=1}^N \sum_{l \in I} \pi_1^i(l) H_3(\pi_{2l}^i)$ بیت

$$L_1 + L_2 = \text{به طوری که می‌شود} \quad \sum_{i=1}^N H_3(\pi^i) = L$$

اعوجاج مشترک بر روی بلاک‌های 2×1 می‌تواند تأثیر متقابل تغییرات در جهت‌های افقی را منعکس کند. برای گنجاندن تأثیر متقابل در جهت‌های عمودی، ردیف‌های فرد تصویر پوشانه را به عنوان یک زیرتصویر و ردیف‌های زوج را به عنوان زیرتصویر دیگر در نظر گرفته می‌شوند. سپس اعوجاج بلاک پیکسل در زیرتصویر دوم با توجه به تغییرات بلاک‌های بالا و پایین به روز می‌شود. همچنین، تعریف اعوجاج در بلاک‌های بزرگ‌تر مانند بلاک‌های 2×2 باعث افزایش مقاومت و بهبود محرمانگی می‌شود.

۵-۲-۵- روش مبتنی بر میدان تصادفی مارکوف گاوسی (GMRF)

در این روش [۳۷]، یک مدل تصادفی گاوسی مارکوف با همسایگی متقابل چهارگانه برای توصیف فعل و انفعالات بین پیکسل‌های محلی تصاویر پوشانه پیشنهاد می‌شود؛ بنابراین مسئله نهان‌نگاری با استفاده از مزایای استقلال مشروط GMRF با هدف به کمینه‌رساندن واگرایی KL از لحاظ یک سری ساختارهای گروهک با ابعاد کم مرتبط با GMRF فرموله می‌شود. در روش پیشنهادی یک طرح بهینه‌سازی تکراری متناوب برای جاسازی مؤثر یک پیام با طول مشخص اعمال می‌شود به طوری که واگرایی KL بین پوشانه و حامل را به کمینه برساند تا امنیت و محرمانگی را بهبود بخشد.

$$X = [X_1, X_2]^T \text{ پوشانه دودویی} \\ \pi = [\pi_1, \pi_2]^T \text{ جاسازی دو پیکسل را با احتمالات} \\ \text{ویرایش کرده و به گروهک حامل } Y = [Y_1, Y_2]^T \text{ تغییر می‌دهد. برای احتمالات تغییر کوچک } \pi \text{ واگرایی KL بین گروهک‌های پوشانه و حامل به خوبی می‌تواند با مؤلفه‌های درجه دوم آن تقریب زده شود:} \\ D(\pi) = D_{KL}(F_P \parallel F_Q^\pi) \approx \frac{1}{2} \pi^T \cdot \nabla^2 D(0) \cdot \pi \quad (31)$$

$$\text{که } F_P \text{ و } F_Q^\pi \text{ به ترتیب بیانگر گروهک‌های دودویی پوشانه و حامل هستند. } \nabla^2 D(0) \text{ مشتق جزئی مرتبه دوم } D \text{ به شرط } \pi = 0 \text{ است. همچنین } \nabla^2 D(\pi) \text{ با ماتریس اطلاعات فشر } I_2(\pi) \text{ در } \pi = 0 \text{ متناسب است.} \\ \nabla^2 D(\pi)|_{\pi=0} = I_2(\pi)|_{\pi=0} / \ln 2 \quad (32)$$

که $I_2(\pi)$ ماتریس اطلاعات فشر نهان‌نگار دودویی است:

Decompose the joint distribution

که توزیع اصلی را با ضرب حاشیه‌های فردی تقریب می‌زند. در استنتاج میدان متوسط، تعاملات در حاشیه‌های فردی گنجانده می‌شوند؛ بنابراین هزینه تغییر پیکسل‌ها به‌صورت جداگانه به‌دست می‌آید که با روش‌های جاسازی عملی مانند STC به‌طور کامل سازگار است.

استفاده از MRF یک راهکار مناسب برای مدل‌سازی بافت تصویر است که تعاملات متقابل بین پیکسل‌ها در یک همسایگی محلی را مدل‌سازی می‌کند؛ از این‌رو، در نهان‌نگاری نیز می‌توان از مدل MRF برای لحاظ نمودن وابستگی بین تغییرات پیکسل‌ها بهره گرفت؛ بنابراین در این روش تابع اعوجاج به‌صورت زیر تعریف می‌شود:

$$D(y) = w_1 \sum_{i=1}^n \rho_i + w_2 \sum_{i=1}^n \sum_{j \in N_i} \rho_{ij} \quad (39)$$

که w_1 و w_2 وزن‌های گروه‌ها و N_i همسایگی محلی پیکسل i را بیان می‌کنند. همچنین، برای سادگی $\rho_i = \rho(y_i)$ در نظر گرفته می‌شود. با لحاظ نمودن وابستگی بین پیکسل‌های مجاور، دیگر نمی‌توان احتمال جاسازی را به‌عنوان ضرب احتمالات حاشیه‌ای فردی فرموله کرد و به‌صورت زیر بیان می‌شود:

$$P(y) = \prod_{i=1}^n p_i \prod_{j \in N_i} \frac{p_{ij}}{p_i p_j} \quad (40)$$

که $p_i \approx \exp(-\rho_i)$ و $p_{ij} \approx \exp(-\rho_{ij})$.

میدان متوسط یک روش بهینه‌سازی است که برای حل مدل‌های MRF، چگالی حاشیه‌ای^۳ هر متغیر را تخمین می‌زند؛ بنابراین به‌جای محاسبه توزیع دقیق $P(y)$ ، توزیع احتمال MRF با یک توزیع فاکتوربندی شده $Q(y)$ تخمین زده می‌شود که می‌تواند به‌صورت ضرب حاشیه‌ای مستقل $Q(y) = \prod_{i=1}^n \prod_{u \in I} q_i^u$ بیان شود. احتمال تغییر q_i^u احتمال اختصاص برچسب $u \in I = \{-1, 0, +1\}$ به پیکسل i را نشان می‌دهد. این احتمال، با هدف کمینه کردن واگرایی KL بین $P(y)$ و $Q(y)$ ، به‌صورت یک معادله به‌روزرسانی تکراری به‌دست می‌آید [19]:

$$q_i^u \propto \exp\{-\lambda(w_1 \rho_i^u + w_2 \sum_{j \in N_i} \sum_v \rho_{ij}^{uv} q_j^v)\} \quad (41)$$

که پارامتر λ از عبارت $-\sum_y Q(y) \ln Q(y) = L$ محاسبه می‌شود.

هزینه‌های اولیه با استفاده از روش HILL محاسبه می‌شوند^۴ $\rho_i^+ = \rho_i^- = \rho_i$ و $\rho_i^0 = 0$.

$$I_2(0) = \begin{bmatrix} I_2(0)_{1,1} & I_2(0)_{1,2} \\ I_2(0)_{2,1} & I_2(0)_{2,2} \end{bmatrix} \quad (33)$$

$$I_2(0)_{k,l} = E \left[\left(\frac{\partial \ln F_2^0(x)}{\partial \pi_k} \cdot \frac{\partial \ln F_2^0(x)}{\partial \pi_l} \right) \right]_{\pi=0}, k \in \{1,2\}, l \in \{1,2\} \quad (34)$$

پس واگرایی KL بین یک جفت گروهک دودویی پوشانه و حامل به‌صورت زیر بازنویسی می‌شود:

$$D_{KL}(F_P \parallel F_Q^\pi) = \frac{(I_2(0)_{1,1} \pi_1^2 + 2I_2(0)_{1,2} \pi_1 \pi_2 + I_2(0)_{2,2} \pi_2^2)}{2 \ln 2} \quad (35)$$

در این روش از نرخ تغییر π به‌عنوان یک معیار مؤثر برای تخصیص پویای گروهک‌های مدل در یک فرآیند بهینه‌سازی بهره می‌گیرد. این امر باعث توزیع جاسازی متراکم‌تری نسبت به مدل‌های گاوسی مستقل مانند MiPOD می‌شود.

روش پیشنهادی تصویر پوشانه را به دو زیرتصویر تقسیم می‌کند $\mathcal{A} \cup \mathcal{B}$ احتمال تغییرات π^B در دامنه $[0, \dots, 0.1]$ مقداردهی می‌شود. سپس

الف) برای $k = 1$ تا $k = 4$ (GMRF) با تخصیص گروهک پویا):

$$\begin{aligned} & 1 - \pi^B \text{ بدون تغییر نگه داشته و } \pi^A \text{ بهینه می‌شود:} \\ & \pi^A = \underset{\pi_{X_s}^A}{\operatorname{argmin}} \{ \sum_{T_s} D_{KL}^{A,T_s} - \lambda [\sum_{X_s} h(\pi_{X_s}^A) - L/2] \} \\ & \lambda^A(k) = \lambda. \end{aligned} \quad (36)$$

$$\begin{aligned} & 2 - \pi^A \text{ بدون تغییر نگه داشته و } \pi^B \text{ بهینه می‌شود:} \\ & \pi^B = \underset{\pi_{X_s}^B}{\operatorname{argmin}} \left\{ \sum_{T_s} D_{KL}^{B,T_s} - \lambda \left[\sum_{X_s} h(\pi_{X_s}^B) - L/2 \right] \right\} \\ & \lambda^B(k) = \lambda. \end{aligned} \quad (37)$$

۳- برای $k \geq 2$:

$$r^A = \lambda^A(k) / \lambda^A(k-1), r^B = \lambda^B(k) / \lambda^B(k-1) \quad (38)$$

اگر $(r^A > 0.98 \& r^B > 0.98)$ ، پس پایان.
ب) هزینه جاسازی پیکسل‌ها بر اساس π^A و π^B و از طریق رابطه $d = \ln(1/\pi - 2)$ محاسبه شده و پیام با استفاده از الگوریتم STC جایگذاری می‌شود.

۶-۲-۵- روش مبتنی بر میدان متوسط

در این روش، وابستگی‌های متقابل بین پیکسل‌های مجاور با استفاده از میدان تصادفی مارکوف جفتی^۱ مدل می‌شود. سپس، برای به‌دست‌آوردن هزینه تغییر هر پیکسل، از استنتاج میدان متوسط^۲ بهره گرفته می‌شود.

¹ Pairwise Markov Random Field

² Mean Field

³ Marginal density



$$\beta_j = \begin{cases} 0 & |d_j - u| = 0 \\ \frac{1}{|d_j - u|} & |d_j - u| \neq 0 \end{cases} \quad (46)$$

برای هر پیکسل: $d_j = x_j - y_j \in \{-1, 0, +1\}$ درحالی که احتمال منتسب به پیکسل‌های ثابت بر اساس تغییر یا عدم تغییر آن‌ها، صفر یا یک در نظر گرفته شده است.

۵-۳- مرور روش‌های مبتنی بر اطلاعات جانبی

در نخستین روش نهان‌نگاری مبتنی بر اطلاعات جانبی [۴۰]، پیام محرمانه با برهم‌زدن روند کمی‌سازی رنگ‌ها و تغییر رنگ هنگام تبدیل یک تصویر با رنگ واقعی به قالب پالت، جاسازی می‌شود. در کمی‌سازی آشفته، پوشانه با فرمت JPEG دوباره فشرده می‌شود تا اطلاعات جانبی به دست آید. اولویت جاسازی با اصلاح ضرایب DCT است که در طول فشرده‌سازی دوم نزدیک به وسط بین‌های کوانتیزاسیون قرار می‌گیرند. از آنجایی که به‌طور معمول آخرین مرحله پردازش کمی‌سازی^۲ است، فرستنده به خطاهای گرد کردن دسترسی دارد و به‌طور معمول از آن‌ها برای تنظیم هزینه‌ها استفاده می‌کند.

۵-۳-۱- استفاده از اطلاعات جانبی در رویکرد مبتنی بر هزینه

در [۴۰]، با فرض اینکه نهان‌نگار به یک پوشانه اولیه (پوشانه غیرکوانتیزه x_{ij}) دسترسی دارد، فرستنده خطای گردکردن را از رابطه $-1/2 \leq e_{ij} = x_{ij} - [x_{ij}]$ محاسبه می‌کند که در آن $[x]$ عملیات گردکردن به نزدیک‌ترین عدد صحیح در محدوده دینامیکی پوشانه را نشان می‌دهد. در صورت عدم جاسازی، نهان‌نگار به‌سادگی تصویر پوشانه را ارسال می‌کند $c_{ij} = [x_{ij}]$. در جاسازی سه‌گانه، هزینه‌های هر دو جهت تغییر بر اساس خطای گردکردن تعدیل می‌شود.

$$\rho_{ij}(\text{sign}(e_{ij})) = (1 - 2|e_{ij}|)\rho_{ij} \quad (47)$$

$$\rho_{ij}(-\text{sign}(e_{ij})) = \rho_{ij} \quad (48)$$

که ρ_{ij} هزینه‌های تعدیل شده است. هزینه‌ها با $1 - |e_{ij}|$ متناسب می‌گردند؛ زیرا وقتی $|e_{ij}| \approx 1/2$ باشد،

سپس، برای هماهنگ‌سازی تعاملات بین پیکسل‌ها یک همسایگی چهارتایی متشکل از پیکسل‌های مجاور عمودی و افقی پیکسل (i, j) تعریف می‌شود: $((i, j)(i-1, j)), ((i, j)(i+1, j)), ((i, j)(i, j-1)), ((i, j)(i, j+1))$. این روش برای دو رویکرد جاسازی متقارن [۳۸] و نامتقارن [۳۹] مورد بررسی قرار گرفته است.

پس با فرض $u, v \in I = \{-1, 0, +1\}$ در رویکرد متقارن (SymMF) گروهک دوتایی به شکل زیر تعریف می‌شود:

$$\rho_{ij}^{uv} = \alpha(u, v) * \tilde{\rho}_{ij} \quad (42)$$

درحالی که:

$$\tilde{\rho}_{ij} = (\rho_i + \rho_j)/2 \quad (43)$$

$$\alpha(u, v) = \begin{cases} 0 & u = v \\ 5 & |u| + |v| = 1 \\ 2 & u \neq v \text{ and } |u| + |v| = 2 \end{cases} \quad (44)$$

در این روش، احتمال ویرایش هر پیکسل به‌طور مستقل به‌روز می‌شود به‌طوری که این احتمال برای هر پیکسل متناسب با احتمالات محاسبه شده برای پیکسل‌های همسایه آن خواهد بود و برای تغییرات مثبت و منفی احتمالات مساوی به‌دست می‌آید.

از آنجایی که احتمال تغییر حاصل از جاسازی در هر پیکسل به تغییرات پیکسل‌های مجاور وابسته است، در جاسازی نامتقارن، این احتمال برای پیام‌های مختلف متفاوت خواهد بود. محاسبه احتمال تغییرات مشروط به پیام، پایداری و امنیت حامل را بهبود می‌بخشد. به این ترتیب، برخی از متغیرها مقادیر ثابت می‌گیرند و زیرمسئله برای متغیرهای باقیمانده مشروط به مقادیر ثابت تقریب زده می‌شود. در این روش، از میدان متوسط ساختاریافته^۱ استفاده شده است تا با تجزیه مسئله به زیرمسائل ساده‌تر، به‌طور کارآمد برخی از انواع وابستگی‌ها کنترل شود. در این راستا، تصویر پوشانه به تعدادی زیرتصویر غیر هم‌پوشان تقسیم می‌شوند. پیام نیز به بخش‌های کوچک‌تر تقسیم شده و هر بخش در یک زیرتصویر جاسازی می‌شود به‌گونه‌ای که مقدار اعوجاج را به حداقل برساند. در رویکرد نامتقارن (AsymMF) گروهک دوتایی به شکل زیر تعریف می‌شود:

$$\rho_{ij}^{uv} = \alpha(u, v) * \beta_j * \tilde{\rho}_{ij}^{uv} \quad (45)$$

درحالی که:

² quantization

¹ structural MF

$$w_{kl} = \begin{cases} 1 - 2|e_{kl}| & \text{when } \text{sign}(e_{kl}v_{kl}) > 0 \\ 1 & \text{otherwise} \end{cases} \quad (50)$$

وزن دهی به تعیین میزان تغییر هر پیکسل در همسایگی متقابل کمک می‌کند. به‌طور خاص، جاسازی تغییرات پیکسل‌ها با $|e_{kl}| \approx 1/2$ دارای $w_{kl} \approx 0$ است درحالی‌که $w_{kl} = 1$ هنگامی است که اطلاعات جانبی و جاسازی واقعی هریک بخواهند پیکسل را در جهت مخالف تغییر دهند.

وقتی عبارت $\mu_{ij}^{(w)} = 0$ برقرار است، تغییرات همسایه بر هزینه‌های تعدیل‌شده با اطلاعات جانبی $\rho_{ij}^{(SI)}(\pm 1)$ تأثیری نمی‌گذارد؛ اما زمانی که $\mu_{ij}^{(w)} \neq 0$ باشد، هزینه‌های جاسازی تعدیل می‌شوند:

$$\rho_{ij}^{(nmSI)}(\text{sign}(\mu_{ij}^{(w)})) = \rho_{ij}^{(SI)}(\text{sign}(\mu_{ij}^{(w)})) \times ((1 - \alpha)(1 - |\mu_{ij}^{(w)}|)^p + \alpha) \quad (51)$$

$\rho_{ij}^{(nmSI)}(-\text{sign}(\mu_{ij}^{(w)})) = \rho_{ij}^{(SI)}(-\text{sign}(\mu_{ij}^{(w)}))$ (52)
که در آن p یک عدد صحیح مثبت است و $0 \leq \alpha \leq 1$ به‌صورت تجربی تعیین می‌شود. جاسازی در زیربخش‌های باقیمانده همان مراحل زیربخش دوم را دنبال می‌کند. این روش "nmSI¹" نامیده می‌شود

۶- نهان‌نگاری مبتنی بر یادگیری تقابلی

نبرد بی‌پایان بین نهان‌نگاری و نهان‌کاوی یک محیط رقابتی ایجاد می‌کند که برای هر دو زمینه مفید است. تکامل روش‌های نهان‌کاوی باعث تمرکز بیشتر جامعه پژوهشی بر روی روش‌های جدید نهان‌نگاری شده است تا بتوانند در برابر مدل‌های نهان‌کاو جدید مقاومت کنند. برای این منظور، محققان توابع اعوجاج پیچیده‌تری را برای به دست آوردن نقشه هزینه جاسازی ارائه می‌دهند. آن‌ها حتی فراتر رفته و برداشت جدیدی از نهان‌نگاری را بر اساس یادگیری عمیق ارائه می‌دهند.

در نهان‌نگاری، فرستنده و گیرنده قصد دارند پیامی را به‌صورت محرمانه مبادله کنند؛ درحالی‌که نهان‌کاو، با عنوان مهاجم، تبادلات بین آن‌ها را مشاهده کرده و بررسی می‌کند که آیا رسانه‌های مبادله‌شده طبیعی بوده یا حامل پیام مخفی هستند. مفهوم بازی بین فرستنده، گیرنده و مهاجم با تئوری بازی مطابقت دارد. هر بازیکن سعی می‌کند استراتژی را دنبال کند که احتمال

یک اغتشاش کوچک x_{ij} می‌تواند باعث گرد شدن x_{ij} به جهت دیگر شود. از سوی دیگر، در حالت $e_{ij} \approx 0$ اطلاعاتی برای اولویت‌دهی به یک جهت وجود ندارد بنابراین هزینه‌ها بدون تغییر می‌مانند.

۵-۳-۲- استفاده از اطلاعات جانبی در رویکرد مبتنی بر مدل

برای طرح‌هایی که قابلیت تشخیص را به‌جای هزینه به کمینه می‌رسانند، نویسندگان در [۴۱] نشان می‌دهند که پوشانه اولیه کمبود مدل پوشانه را جبران می‌کند. در این روش اطلاعات فیشر نهان‌نگاری باید بر اساس $(1 - 2|e_{ij}|)^2$ تنظیم شود. در نسخه مبتنی بر اطلاعات جانبی برای روش‌های مبتنی بر مدل، ابتدا احتمالات تغییر جاسازی متقارن π_{ij} بر اساس کمینه‌سازی واگرایی KL بین پوشانه و حامل محاسبه می‌شود. سپس فرستنده با استفاده از معکوس معادله (۹) احتمالات را به هزینه‌های تغییر جاسازی تبدیل می‌کند و در نهایت با ترکیب اطلاعات جانبی، هزینه‌ها را تعدیل می‌کند.

۵-۳-۳- ترکیب اطلاعات جانبی و تعاملات بین پیکسل‌ها

در این بخش، ترکیب دو رویکرد بهره‌گیری از اطلاعات جانبی [۴۲] و خوشه‌بندی جهت تغییرات [۳۴] مورد بررسی قرار می‌گیرد. باید توجه داشت که نتایج ارائه‌شده توسط اطلاعات جانبی و تغییرات مجاور گاهی ممکن است در تضاد باشند و جهت تغییرات همسایه به جهتی متفاوت از اطلاعات جانبی اشاره کند. در این راستا، در [۴۳] ابتدا تمام هزینه‌ها بر اساس اطلاعات جانبی تعدیل می‌شود. عملیات جاسازی در زیربخش‌های غیر هم‌پوشان انجام می‌شود و در همین حین هزینه‌های تغییرات در جهات مختلف توسط یک عامل ضربی تنظیم می‌شود که به‌طور غیرخطی به میانگین محلی تغییرات وابسته به خطاهای گرد کردن آن‌ها بستگی دارد.

در اولین زیربخش، جاسازی بر اساس هزینه‌های به‌دست‌آمده از اطلاعات جانبی $\rho_{ij}^{(SI)}(\pm 1)$ انجام می‌شود. در زیربخش دوم، برای هر پیکسل، ابتدا میانگین وزن‌دار تغییرات جاسازی $v_{kl} \in \{-1, 0, 1\}$ در یک همسایگی متقاطع محاسبه می‌شود:

$$\mu_{ij}^{(w)} = \frac{1}{4} \sum_{(k,l) \in C_{ij}} w_{kl} v_{kl} \quad (49)$$

که وزن‌ها از رابطه زیر به دست می‌آیند:

¹ Neighborhood Modulated Side-Informed

شبکه، می‌توان با استفاده از p_g نمونه‌ها را تا حد امکان واقعی تولید کرد. تابع هدف مدل GAN به صورت زیر بیان می‌شود:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (53)$$

برای برنده شدن در این بازی، دو بازیکن (D, G) باید به طور مداوم پارامترهای خود را بهینه کنند تا ظرفیت تولید را افزایش دهند. در حالت بهینه، معیار توقف فرآیند بهینه‌سازی دستیابی به تعادل نش بین دو بازیکن است. وقتی هیچ‌یک از بازیکنان نتوانند وزن‌های خود را بدون تغییر در پارامترهای بازیکن دیگر کاهش دهند، تعادل نش حاصل می‌شود. در این حالت، توزیع آموخته شده و توزیع اصلی بهینه هم‌گرایی $p_g \approx p_{data}$ خود را به دست می‌آورند.

۶-۲- نمونه‌های تقابلی

در سال ۲۰۱۳ Szegedy اعلام کرد که با وجود کارایی بالای شبکه‌های عصبی عمیق و مقاومت آن‌ها نسبت به آشفتگی‌های تصادفی، این شبکه‌ها نسبت به حملات تقابلی آسیب‌پذیر هستند [۵۶]. نمونه‌های تقابلی از طریق اغتشاشات غیر تصادفی جزئی در داده‌های ورودی، با هدف بهینه‌کردن خطای طبقه‌بندی به وجود می‌آیند. این اغتشاشات را می‌توان بر اساس معکوس جهت‌گردان شبکه هدف به دست آورد. نمونه‌های تقابلی شبکه را گمراه می‌کنند تا با اطمینان بالایی خروجی‌های نادرست تولید کند؛ بنابراین اعمال آشفتگی نمونه‌های تقابلی برای فریب طبقه‌بند، حمله تقابلی نامیده می‌شود؛ اما این رویکرد برخلاف GAN که یک فرآیند بازی تکراری و پویا بین مولد و تمایزدهنده برای رسیدن به تعادل نش است، یک فرآیند پویا نیست. در این رویکرد، مدل از دو شبکه نهان‌نگار و نهان‌کاو تشکیل می‌شود که شبکه نهان‌کاو از قبل آموزش دیده و در این مدل به عنوان منتقد عمل می‌کند. شبکه نهان‌نگار در این مدل تصویر حامل را در قالب نمونه تقابلی به گونه‌ای تولید می‌کند که نهان‌کاو را به سمت نتیجه اشتباه هدایت کند. نمونه‌های تقابلی ماهیت مشترک مدل‌های مختلف را تحت تأثیر قرار می‌دهد. از این رو، مهاجمان می‌توانند بدون اطلاع از ساختار داخلی مدل هدف، مدل را مورد حمله قرار دهند؛ بنابراین می‌توان از یک مدل شبکه آموزش دیده به عنوان مولد حملات تقابلی استفاده کرد.

برنده شدن خود را به بیشینه برساند و هرگونه تغییر استراتژی یک بازیکن، باعث یک ضد حمله از طرف بازیکنان دیگر می‌شود که می‌تواند منجر به افزایش سود آن‌ها شود. از این رو، این بازی به عنوان یک مسئله بهینه‌سازی min-max بیان می‌شود. پاسخ بهینه، در صورت وجود، پاسخ در تعادل نش^۱ نامیده می‌شود.

در سال ۲۰۱۴، گودفیلو و همکارانش [۵۵] از مفاهیم شبکه عصبی برای شبیه‌سازی یک بازی استفاده کردند که شبکه مولد تقابلی (GAN) نامیده شد. GAN یک شبکه یادگیری عمیق بدون نظارت^۲ است که یک شبکه مولد^۳ و یک شبکه تمایزدهنده^۴ را در یک بازی minimax دو عامله آموزش می‌دهد. رویکرد GAN در نهان‌نگاری می‌تواند برای بهبود نتیجه روش‌های سنتی موجود و یا ارائه یک مدل یکپارچه برای نهان‌نگاری مورد استفاده قرار گیرد. در رویکرد نهان‌نگاری، تمایزدهنده یک شبکه نهان‌کاو در نظر گرفته می‌شود و مولد یک شبکه نهان‌نگار خواهد بود که سعی می‌کند پیام را به گونه‌ای جاسازی کند که تصویر تولیدشده تا حد امکان واقعی بوده و نهان‌کاو امکان تمیز آن از تصاویر اصلی را نداشته نباشد. در این سناریو، شبکه‌های نهان‌نگار و شبکه نهان‌کاو به طور مکرر دانش خود را در مورد شبکه رقیب به روز می‌کنند.

۶-۱- شبکه مولد تقابلی

شبکه مولد تقابلی [55] (GAN) به عنوان یک مدل مولد قدرتمند حاوی دو شبکه عصبی مولد G و تمایزدهنده D است که در یک بازی minimax دو عامله با یکدیگر رقابت می‌کنند. در مدل GAN، هدف مولد تولید تصاویری با توزیع احتمال مشابه تصاویر واقعی است. شبکه مولد از نتایج ارزیابی شبکه رقیب بر اساس مقایسه تصاویر تولید شده (جعلی) با تصاویر واقعی، به منظور بهبود کیفیت تصاویر تولیدی استفاده می‌کند.

مولد $G(z, \theta)$ یک شبکه عصبی با پارامترهای θ است که برای گمراه کردن تمایزدهنده آموزش می‌بیند. خروجی مولد یک نمونه توزیع از فضای داده $G(z, \theta) \sim p_g$ است. با توجه به متغیر ورودی z از توزیع $p_z(z)$ ، شبکه مولد p_g آموزش می‌بیند تا تصاویری تولید کند که تا حد ممکن به توزیع داده‌های آموزش نزدیک باشند، درحالی‌که تمایزدهنده برای تشخیص نمونه‌های تولید شده از داده‌های واقعی آموزش می‌بیند. پس از آموزش

¹ Nash equilibrium

² Unsupervised

³ Generator

⁴ Discriminator

است. این چارچوب در شبیه‌سازی بازی تقابلی از سه عامل استفاده می‌کند: فرستنده (شبکه نهان‌نگار)، گیرنده (شبکه استخراج‌کننده) و مهاجم (شبکه نهان‌کاو).

۷- مرور روش‌های نهان‌نگاری مبتنی بر یادگیری تقابلی

۷-۱- روش‌های مبتنی بر تخمین ماتریس احتمال

در رویکرد تخمین ماتریس احتمال، هدف شبکه مولد در GAN (نهان‌نگار) محاسبه نقشه احتمال جاسازی بهینه است، در حالی که شبکه تمایزدهنده (نهان‌کاو) عمدتاً برای تشخیص تصویر پوشانه از تصویر حامل متناظر با آن است. در ادامه چند روش ارائه شده با این رویکرد معرفی می‌شود.

۷-۱-۱- یادگیری اعوجاج نهان‌نگاری به‌صورت خودکار

در این راستا، تانگ و همکاران [۶۲] چارچوب یادگیری اعوجاج نهان‌نگاری به‌صورت خودکار (ASDL-GAN) را پیشنهاد دادند تا مکان‌های مناسب برای جاسازی را با کمک یادگیری نقشه احتمال شناسایی کند. شبیه‌ساز برای جاسازی پیام نیز به‌صورت یک تابع فعال‌سازی به‌صورت زیر طراحی می‌شود که p_{ij} احتمال تغییر و n_{ij} یک عدد تصادفی از توزیع یکنواخت در بازه $[0, 1]$ را نشان می‌دهند.

$$m'_{ij} = \begin{cases} -1 & n_{ij} < p_{ij}/2 \\ 1 & n_{ij} > 1 - p_{ij}/2 \\ 0 & \text{otherwise} \end{cases} \quad (54)$$

سپس، یانگ و همکاران مدل [63] UT-GAN (شکل ۵) را برای بهبود مدل ASDL-GAN معرفی کردند که به دلیل عدم انتقال گرادینان در طول باز انتشار، برای شبیه‌سازی جاسازی پیام از تابع فعال‌سازی Tanh بهره می‌گیرد و بنابراین از سرعت و دقت بالاتری برخوردار است.

$$m'_{ij} = -0.5 \times \tanh(\lambda(p_{ij} - 2 \times n_{ij})) + 0.5 \times \tanh(\lambda(p_{ij} - 2 \times (1 - n_{ij})))$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (56)$$

که λ فاکتور مقیاس است.

با وجود این‌که نمونه‌های تقابلی به‌طور معمول به‌عنوان عیب شبکه‌های عصبی شناخته می‌شوند، اما گاهی اغتشاشات افزوده‌شده حاوی اطلاعات معناداری هستند؛ به همین دلیل می‌توان از این رویکرد برای بهبود عملکرد نهان‌نگاری استفاده کرد. در این راستا، هزینه‌های تغییر به‌دست‌آمده از روش‌های سنتی موجود، بر اساس گرادینان شبکه نهان‌کاو اصلاح می‌شوند [۵۷-۶۱].

۶-۳- ویرایش پوشانه بر اساس یادگیری تقابلی

در ادامه به بررسی روش‌های مبتنی بر ویرایش پوشانه براساس یادگیری تقابلی پرداخته می‌شود که تاکنون به چند روش انجام گرفته است:

- تخمین ماتریس احتمال ویرایش پوشانه
- نهان‌نگاری براساس نمونه‌های تقابلی
- نهان‌نگاری براساس بازی تقابلی سه‌عامله

۶-۳-۱- تخمین ماتریس احتمال

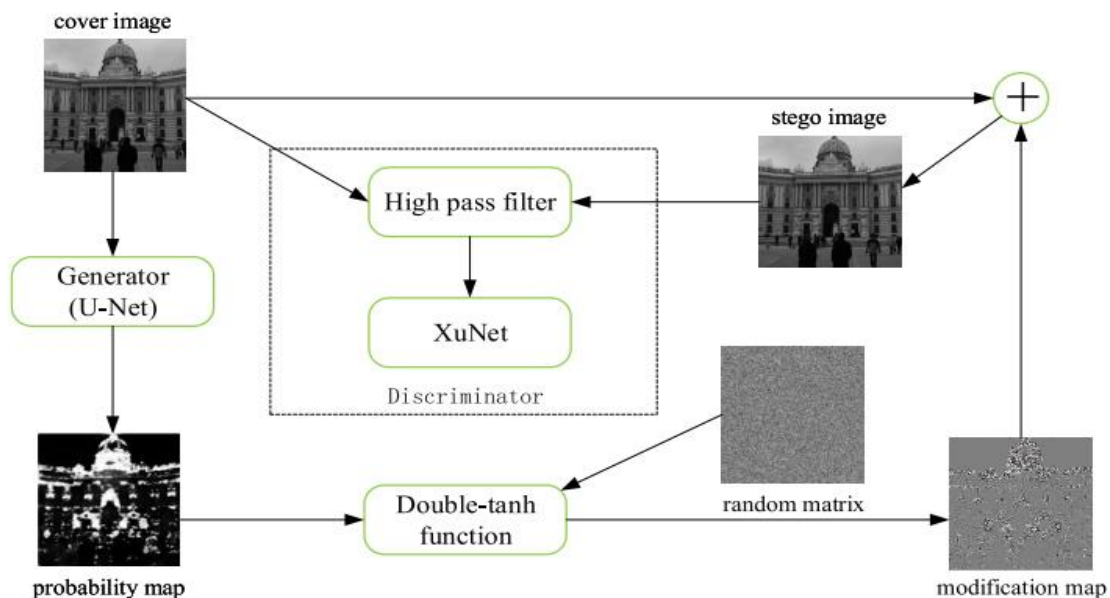
در این روش با استفاده از GAN، ماتریس احتمال ویرایش تصویر پوشانه با هدف به کمینه‌رساندن اعوجاج به دست می‌آید و سپس با استفاده از روش‌های کدگذاری مناسب، پیام در پوشانه جاسازی می‌شود.

۶-۳-۲- نهان‌نگاری براساس نمونه‌های تقابلی

ایده استفاده از نمونه‌های تقابلی در نهان‌نگاری به‌تازگی توجه برخی پژوهش‌گران این حوزه را به خود جلب کرده است. رویکردهایی که در این خانواده قرار می‌گیرند، اغلب از مفهوم شبیه‌سازی بازی با دو عامل (فرستنده و مهاجم) استفاده می‌کنند تا نقشه هزینه را محاسبه کنند و عملیات جاسازی و استخراج پیام را به روش‌های کدگذاری مانند STC می‌سپارند.

۶-۳-۳- نهان‌نگاری براساس بازی تقابلی سه‌عامله

روش‌های مبتنی بر اعوجاج، بر توسعه تابع اعوجاج متمرکز هستند و وظیفه جاسازی پیام را به برنامه‌های کدگذاری عملی می‌سپارند. با توجه به تأثیر فرآیند (۵۵) جاسازی بر کیفیت تصویر، چارچوب نهان‌نگاری یک‌پارچه با الهام از مدل‌های یادگیری عمیق ارائه می‌شود که شامل یک مدل رمزنگار-رمزگشا برای یادگیری هم‌زمان جاسازی و استخراج پیام‌های محرمانه



شکل ۵: معماری قالب UT-GAN [63]
Figure 5: UT-GAN architecture [63]

باید ظرفیت واقعی جاسازی را علاوه بر آموزش تقابلی بین تمایزدهنده و مولد در نظر بگیرد. در نتیجه، تابع زیان برای مولد از دو بخش زیان تقابلی GL_G^1 و زیان آنتروپی GL_G^2 تشکیل می‌شود که در آن زیان تقابلی تشخیص تصاویر نهان‌نگاری را دشوارتر می‌کند و می‌تواند به عنوان معکوس زیان نهان‌کاو در نظر گرفته شود، درحالی‌که زیان آنتروپی تضمین می‌کند که تصویر حامل با نرخ تقریبی Q جاسازی شده است. بر این اساس

$$GL_G^1 = -L_D \quad (58)$$

$$GL_G^2 = (C - H \times W \times Q)^2 \quad (59)$$

که $H \times W$ ابعاد تصویر و C ظرفیت واقعی جاسازی را مشخص می‌کنند.

در نهایت، تابع زیان کل نهان‌نگار L_G با ترکیب تابع زیان تقابلی GL_G^1 و ازدست‌دادن آنتروپی GL_G^2 تعریف می‌شود.

$$L_G = \alpha \times GL_G^1 + \beta \times GL_G^2 \quad (60)$$

که در آن α و β دو پارامتر تنظیم شده هستند که می‌توانند به عنوان مقدار تجربی $\alpha = 1$ و $\beta = 10^{-7}$ تنظیم شوند.

۷-۱-۲- نهان‌نگاری با استفاده از راهکار بازخورد متقابل^۱

روش‌های نهان‌نگاری مبتنی بر GAN اغلب از شبکه‌های عصبی کانولوشنال (CNN) خطی استفاده می‌کنند. ارتباط خطی بین لایه‌های ارتباط^۲ و گسترش^۳ باعث می‌شود که اطلاعات لایه ارتباط به‌طور کامل به لایه گسترش برگشت داده نشوند و در نتیجه به دلیل آموزش ناکافی شبکه، محرمانگی و امنیت نهان‌نگار کاهش می‌یابد. در این راستا، در [۶۴] یک مدل نهان‌نگاری مبتنی بر GAN ارائه می‌شود که با ایجاد چند کانال بازخورد متقابل در شبکه نهان‌نگار، اجازه می‌دهد اطلاعات به‌طور مستقیم از طریق این کانال‌های بازخورد به لایه گسترش ارسال شود. از آنجایی که در این روش اطلاعات دقیق لایه‌های عمیق را می‌توان به‌طور موثر با مراجعه به اطلاعات تصویر پوشانه دریافت کرد، نهان‌نگار در نهایت می‌تواند یک نقشه احتمال مناسب برای جاسازی با امنیت بالا بیاموزد (شکل ۶).

تابع زیان نهان‌کاو به صورت زیر تعریف می‌شود:

$$L_D = -\sum_{i=1}^2 y_i' \log(y_i) \quad (57)$$

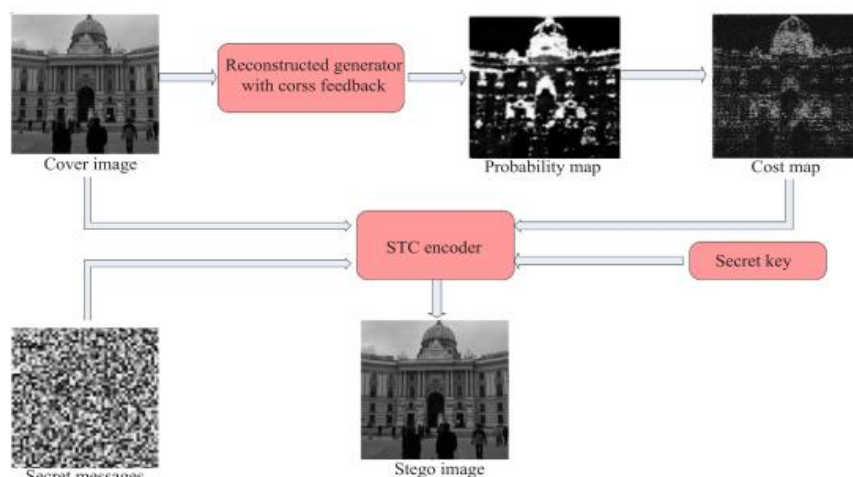
که y_i خروجی لایه softmax و y_i' برچسب واقعی است.

از آنجایی که خروجی شبکه نهان‌نگار، نقشه احتمال جاسازی براساس یک ظرفیت مشخص است، تابع زیان

¹ cross feedback

² contraction

³ expansion



(شکل - ۶): معماری قالب CF-GAN [64]

Figure 6: CF-GAN architecture [64]

فریب نهان‌کاو است. برای این منظور همان‌طور که در شکل (۷) نشان داده شده است، پس از مقداردهی اولیه β ، در مرحله نخست بخشی از پیام محرمانه با استفاده از یک روش نهان‌نگاری سازگار با محتوای سنتی در گروه مشترک جایگذاری می‌شود تا تصویر Z_c به دست آید (C تصویر پوشانه است)؛ سپس در مرحله دوم، باقیمانده پیام با کمک طرح جاسازی تقابلی در گروه قابل تنظیم جاسازی می‌شود، بنابراین تصویر Z تولید می‌شود.

پیکسل‌های قابل تنظیم در راستای فریب نهان‌کاو هدف ویرایش می‌شوند به‌گونه‌ای که نهان‌کاو بر اساس هزینه‌های جاسازی عناصر قابل تنظیم $(+q, -q)$ تصویر حامل به‌دست‌آمده را به‌عنوان پوشانه تشخیص دهد. اگر تصویر حاصل Z نهان‌کاو را با موفقیت همراه کند، به‌عنوان تصویر نهایی در نظر گرفته و بنابراین تکرار متوقف می‌شود. در غیر این صورت، تعداد پیکسل‌های قابل تنظیم $\beta = \beta + \Delta\beta$ را افزایش داده و مراحل بالا تکرار می‌شود. برای این کار از نهان‌کاو مبتنی بر یادگیری عمیق بهره گرفته شده و از مقادیر گرادینان تابع زیان آن برای هدایت تغییرات پیکسل‌های گروه قابل تنظیم استفاده می‌شود. به‌منظور تطبیق بهتر با نهان‌کاو، هدف و بهبود عملکرد امنیتی در مقابل دیگر نهان‌کاوها، تلاش می‌شود تعداد پیکسل‌های قابل تنظیم به کمینه برسد.

۷-۲-۲- نهان‌نگاری براساس نمونه‌های تقابلی

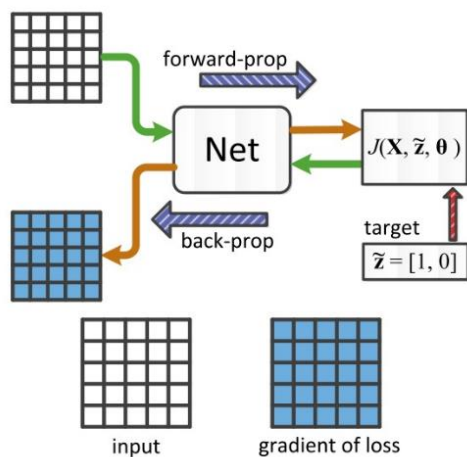
شبکه عصبی در اصل یک تابع نگاشت است که می‌تواند با مؤلفه‌ها و معماری‌های مختلف هر تابعی را شبیه‌سازی کند. در روش AENAdversarial ENhancing [59]

۷-۲-۱- روش‌های مبتنی بر نمونه‌های تقابلی

از نظر تاریخی، ایده نمونه‌های تقابلی در نهان‌نگاری یک ایده جدید نیست. اولین رویکردهای تکراری تقابلی با [65] MOD و [66] ASO ارائه شدند؛ اما این رویکردها پویا نبوده و از یادگیری تقابلی استفاده نمی‌کنند. یک کاربرد نمونه تقابلی، کنترل تغییرات پیکسل‌ها است که مشابه ایده همگام‌سازی تغییرات است که در آن تأثیر متقابل جاسازی در پیکسل‌های مجاور موردتوجه قرار می‌گیرد و با این فرض هزینه پیکسل‌ها به‌روزرسانی می‌شوند. در همگام‌سازی تغییرات، هزینه تغییر پیکسل‌ها با ضرب یا تقسیم در یک عدد ثابت اصلاح می‌شوند؛ اما در روش نمونه تقابلی از مقدار و جهت گرادینان برای بروز رسانی هزینه‌ها استفاده می‌شود. در این راستا، چند روش نهان‌نگاری مبتنی بر نمونه‌های تقابلی ارائه شده است که در ادامه به‌اختصار هر یک معرفی می‌شوند.

۷-۲-۱-۱- جاسازی تقابلی برای نهان‌نگاری

در روش ADV-EMB (Adversarial Embedding) [58] فرستنده به گرادینان تابع زیان نهان‌کاو دسترسی دارد و نقشه هزینه‌ها را با توجه به گرادینان بازگشتی به‌دست‌آمده از نهان‌کاو به‌روز می‌کند؛ بنابراین، جهت تغییر مقدار هزینه بر اساس معکوس جهت گرادینان مشخص خواهد شد. قبل از شروع بازی، نقشه هزینه با کمک یک روش سنتی مقداردهی می‌شود. در این روش، پیکسل‌های تصویر پوشانه به‌صورت تصادفی به دو گروه مشترک و قابل تنظیم تقسیم می‌شوند. جاسازی پیام نیز در دو مرحله انجام می‌شود. هدف این روش به کمینه‌رساندن تعداد پیکسل‌های قابل تنظیم β برای



(شکل - ۸): نهان نگاری مکانی تطبیقی بر اساس

نمونه‌های تقابلی [59]

Figure 8: Adaptive spatial steganography based on adversarial samples [59]

برای فریب شبکه و دریافت پاسخ دسته‌بندی اشتباه $\bar{y}$ ، تابع زیان $J(X, \bar{y}, \theta)$ فاصله ورودی تا $\bar{y}$ را اندازه‌گیری می‌کند. از آنجایی که تمام مؤلفه‌های ورودی در خروجی به‌دست‌آمده مشارکت دارند، آشفتگی از مشتق تابع زیان $\nabla_X J(X, \bar{y}, \theta)$ به دست خواهد آمد. با توجه به اینکه مقدار $\bar{y}$ بسیار کوچک است، به‌طورمعمول از علامت این تابع برای تعیین آشفتگی استفاده می‌شود:

$$\eta = \text{sign}(\nabla_X J(X, \bar{y}, \theta)) \quad (61)$$

با جایگزینی تابع نگاشت به‌جای تابع زیان، هزینه تغییر پیکسل‌ها به صورت زیر ویرایش می‌شود:

$$\begin{cases} (\rho_{ij}^*)' = A(G_{ij}) \times \rho_{ij}^* & \text{if } * = \text{sign}(G_{ij}) \\ (\rho_{ij}^*)' = \rho_{ij}^* & \text{if } * \neq \text{sign}(G_{ij}) \end{cases} \quad (62)$$

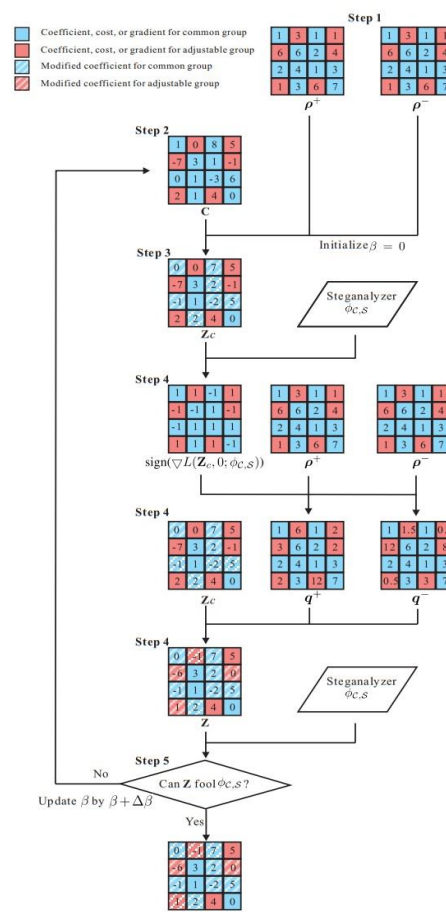
که $G = \nabla_X F_1(X)$ ، $* \in \{+, -\}$ و $A(G_{ij})$ یک پارامتر مقیاس‌گذاری است که بر اساس $0 < \|G_{ij}\| < 1$ به دست می‌آید.

۷-۲-۳- نهان نگاری تقابلی براساس تولید و انتخاب حامل

در روش [۶۰] یک روش جدید نهان نگاری با بهره‌گیری از ایده نمونه‌های تقابلی ارائه می‌شود. چارچوب پیشنهادی شامل سه مرحله است: پیش‌آموزش شبکه نهان‌کاو، تولید حامل و انتخاب حامل نهایی. هدف اغلب روش‌های موجود تولید یک حامل است که نهان‌کاو هدف را فریب دهد. برای بهبود تنوع حامل، روش پیشنهادی ابتدا چند حامل کاندید ایجاد می‌کند و از بین آن‌ها حامل بهینه را انتخاب می‌کند (شکل ۹).

تابع نگاشت به‌صورت $F: R^{m \times n} \rightarrow \{0, 1\}^k$ معرفی می‌شود. ورودی شبکه عصبی کانولوشنی^۱ (CNN)، یک تصویر $m \times n$ و خروجی یک بردار دسته "one-hot" با تعداد ابعاد k است. در مسئله نهان‌کاو سناریوی دسته‌بندی دودویی است $k = 2$ بنابراین خروجی شبکه عصبی نهان‌کاو احتمال پوشانه‌بودن و حامل‌بودن است که به ترتیب با p_1 و p_2 نشان داده می‌شوند. به‌طورمعمول لایه آخر یک شبکه عصبی لایه softmax است؛ بنابراین $p_1 = 1 - p_2$.

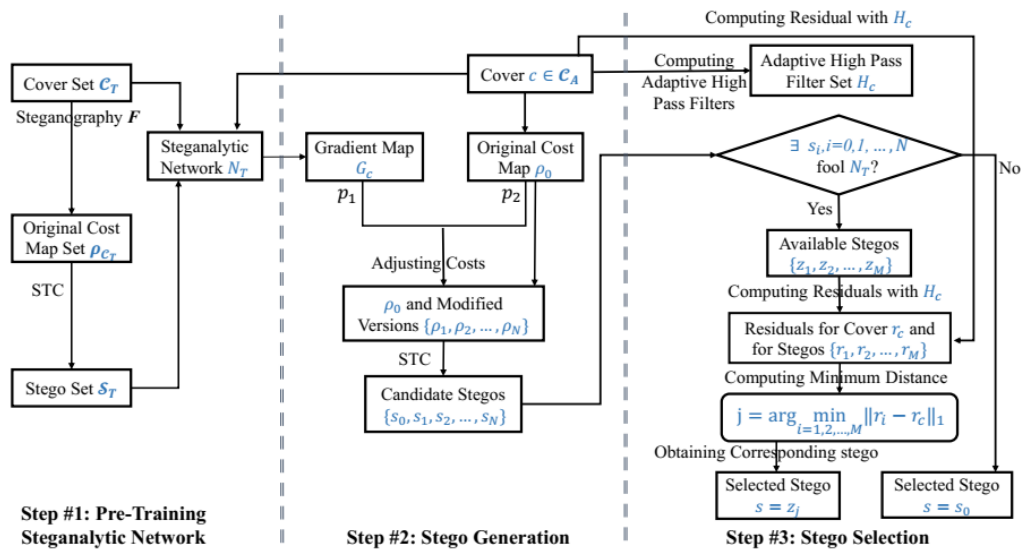
با در نظر گرفتن X به‌عنوان ورودی، تابع نگاشت شبکه عصبی می‌تواند به‌صورت $F_1(X)$ بیان شود که اندیس یک بیان‌گر احتمال دسته یک است؛ بنابراین $p_1 = F_1(X)$ ، $p_2 = 1 - F_1(X)$. در مرحله آموزش مدل شبکه عصبی، هدف بهینه‌سازی تابع زیان $J(X, y, \theta)$ است که نشان‌دهنده خطای بین خروجی مدل و برچسب اصلی داده ورودی (خروجی مورد انتظار) است. شکل (۸) فرایند انتشار و پس‌انتشار را نشان می‌دهد که در انتشار روبه‌جلو $J(X, y, \theta)$ و در پس‌انتشار $\nabla_X J(X, y, \theta)$ محاسبه می‌شود.



(شکل - ۷): مدل ADV-EMB [58]

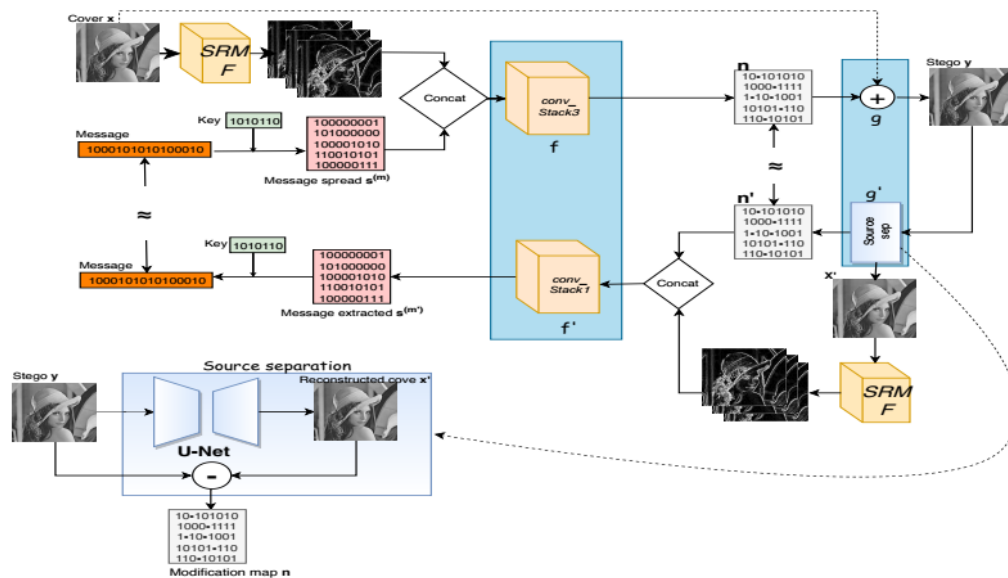
Figure 7: ADV-EMB model[58]

¹ Convolutional Neural Network



(شکل - ۹): مدل تولید و انتخاب حامل [60]

Figure 9: The Generation & Selection method [60]



(شکل - ۱۰): مدل بازی سه عامله - یدرودج [68]

Figure 10: Three-player game model-Yedroudj [68]

نهان‌نگاری F به دست می‌آید؛ سپس نقشه گرادیان G_c با استفاده از یک الگوریتم انتشار برگشتی به شرح زیر محاسبه می‌شود:

$$G_c = \nabla_c L(c, t; N_T) \quad (64)$$

$$L(c, t; N_T) = -t \log(N_T(c)) - (1-t) \log(1 - N_T(c)) \quad (65)$$

که در آن پارامتر t بیانگر حامل یا پوشانه بودن تصویر و تابع زبان آنتروپی متقاطع باینری^۱ است. سپس، هزینه‌های جاسازی به صورت زیر اصلاح می‌شوند:

¹ binary cross entropy

در مرحله نخست، برای مجموعه تصاویر پوشانه $\mathcal{C}_T$ ، با استفاده از یک روش نهان‌نگاری موجود F (HILL و UNIWARD) و هزینه‌های ρ_0 ، مجموعه متناظر تصاویر حامل S_T تولید می‌شود.

$$s_0 = h_{emb}(c, m, \rho_0^+, \rho_0^-) \quad (63)$$

که $\rho_0^+ = \rho_0^- = \rho_0$. سپس مجموعه تصاویر $\mathcal{C}_T$ و S_T به دو مجموعه آموزشی و اعتبارسنجی تقسیم شده و نهان‌کاو N_T آموزش داده می‌شود.

در مرحله دوم، برای هر تصویر پوشانه $c \in \mathcal{C}_A$ که ابتدا تابع هزینه اصلی ρ_0 با روش $\mathcal{C}_A \cap \mathcal{C}_T = \emptyset$

محرمانگی را بهبود می‌بخشد (شکل ۱۰). تابع زیان به صورت زیر محاسبه می‌شود:

$$\mathcal{L} = \lambda_A \cdot (\text{dist}(x, y) - \beta) + \lambda_B \cdot \mathcal{L}_B - \lambda_E \cdot \mathcal{L}_E \quad (71)$$

$$\text{dist}(x, y) = \left(\sum_{i=1}^w \sum_{j=1}^h (|n_{ij}|) \right) / w \cdot h \quad (72)$$

که $\lambda_A, \lambda_B, \lambda_E \in [0, 1]$. پارمتر β به معنی گسسته‌سازی شبکه جاسازی، تعداد پیکسل‌هایی از تصویر پوشانه را که فرستنده مجاز به تغییر آن‌هاست کنترل می‌کند. $\text{dist}(x, y)$ فاصله میانگین مربع خطا (MSE) بین تصاویر پوشانه اصلی و بازیابی‌شده، $\mathcal{L}_B$ فاصله میانگین مربع خطا (MSE) بین پیام اصلی و بازیابی‌شده و همچنین $\mathcal{L}_E$ فاصله cross-entropy بین برچسب ورودی و پیش‌بینی نهان‌کاو را محاسبه می‌کند.

۷-۳-۲- نهان‌نگاری با استفاده از شبکه مولد تقابلی سازگار با چرخه^۱

مدل [۶۹] HCGAN از شبکه باقیمانده^۲ و شبکه مولد تقابلی سازگار با چرخه بهره گرفته و بدین ترتیب موفق شده تصاویری با کیفیت بهتر و محرمانگی بیشتر تولید کند. در این روش ورودی سامانه یک تصویر پوشانه C با ابعاد $h \times w \times c$ (c شماره کانال است) و یک پیام محرمانه M با ابعاد $1 \times l$ است. پیام به اندازه ابعاد تصویر تکرار و به انتهای تصویر پوشانه اضافه می‌شود، اندازه بردار حاصل $h \times w \times (c + l)$ خواهد بود؛ سپس این بردار به‌عنوان ورودی به شبکه نهان‌نگار داده می‌شود. خروجی این شبکه نقشه تغییرات است که با اضافه‌شدن به پوشانه تصویر حامل C' را تولید می‌کند که به‌عنوان ورودی به شبکه‌های گیرنده و نهان‌کاو داده می‌شود (شکل ۱۱). تفاوت بین C و C' با کمک فاصله L_2 محاسبه می‌شود. گیرنده پیام مخفی M' را استخراج می‌کند. فاصله L_2 برای محاسبه تفاوت M و M' استفاده می‌شود (L_H) و L_I شبکه نهان‌نگار معکوس است. نهان‌کاو برچسب پوشانه یا حامل را برای تصویر ورودی مشخص می‌کند. زیان Sigmoid cross entropy برای خطای این شبکه استفاده می‌شود (L_{D1}).

$$L_I(\theta_s, \theta_i; C, C') = \lambda d(C, I(\theta_i; C')) + (1 - \lambda) L_{D_2}(\theta_{d2}; C'') = \lambda d(C, I(\theta_i; C')) - (1 - \lambda) \log(1 - D_2(\theta_{d2}; I(\theta_i; C'))) \quad (73)$$

$$L_S(\theta_s; C, M) = \lambda_s d(C, C') + \lambda_h L_H + \lambda_i L_I + \lambda_{d1} L_{D1}(\theta_{d1}; C') \quad (74)$$

¹ CycleGAN
² residual

$$\begin{aligned} (66) \quad \rho_k^+(i, j) &= \begin{cases} \rho_0^+(i, j) * \alpha, & \text{if } G_c(i, j) > 0, \\ \rho_0^+(i, j), & \text{otherwise,} \end{cases} \\ (67) \quad \rho_k^-(i, j) &= \begin{cases} \rho_0^-(i, j) * \alpha, & \text{if } G_c(i, j) < 0, \\ \rho_0^-(i, j), & \text{otherwise,} \end{cases} \end{aligned}$$

که در آن α نشان دهنده دامنه تقابلی است ($\alpha > 1$). در نهایت، تصاویر حامل کاندید با استفاده از شبیه‌ساز STC تولید می‌شوند:

$$s_k = h_{emb}(c, m, \rho_k^+, \rho_k^-), 0 \leq k \leq N \quad (68)$$

در مرحله آخر، حامل نهایی از بین کاندیدها انتخاب می‌شود. در این راستا، ابتدا یک سری از حامل‌ها براساس میزان توانایی در فریب‌دادن نهان‌کاو N_T انتخاب می‌شوند.

$$Z = \{z \mid z \in \{s_0, s_1, \dots, s_N\} \& N_T(z) < 0.5\} \quad (69)$$

سپس، برای انتخاب حامل بهینه از بین حامل‌هایی که توانسته‌اند نهان‌کاو هدف را فریب دهند، فاصله باقیمانده بین حامل‌ها $r_i, i = 1, 2, \dots, M$ و پوشانه r_c با استفاده از مجموعه فیلترهای تطبیقی بالاگذر H_c محاسبه می‌شود. هر چه این فاصله کمتر باشد، تفاوت آماری بین آن‌ها نیز کمتر و بنابراین امنیت حامل بیشتر خواهد بود.

$$s = z_j, j = \arg \min_{i=1, 2, \dots, M} \|r_i - r_c\|_1 \quad (70)$$

۷-۳-۳- روش‌های مبتنی بر بازی سه‌عامله

در این بازی رقابتی، آموزش بین شبکه‌های نهان‌نگار/ استخراج‌کننده با هدف هم‌گرایی بازیابی پیام از یک‌طرف و شبکه‌های نهان‌نگار/ نهان‌کاو برای ایجاد یک تصویر حامل امن و غیرقابل‌شناسایی از طرف دیگر به‌صورت متناوب انجام می‌گیرد. در ابتدای آموزش، مدل هنوز یاد نگرفته است که برای جاسازی پیام، تغییری مناسب در تصویر پوشانه ایجاد کند؛ بنابراین تصاویر پوشانه و حامل مشابه یکدیگر خواهند بود و شبکه نهان‌کاو نمی‌تواند تصویر پوشانه و حامل را از هم تمیز دهد. استخراج‌کننده نیز به‌طور تصادفی پیام را حدس می‌زند. با ادامه آموزش، مدل سعی می‌کند بین ایجاد تصویر حامل نامحسوس برای فریب نهان‌کاو و استخراج دقیق پیام مخفی تعادل برقرار کند. لازم به ذکر است که این رویکرد نیازمند تعداد تکرارهای زیادی برای رسیدن به تعادل است.

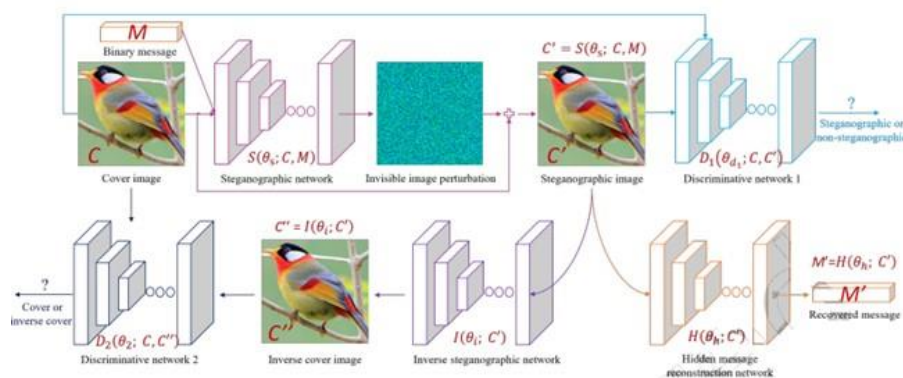
۷-۳-۱- نهان‌نگاری با استفاده از بازی سه

بازیکنه

در این راستا، یدرودج و همکاران [۶۸] معماری ارائه کردند که علاوه بر لحاظ نمودن قدرت نویز حامل، با تعامل بهتر بین شبکه‌های جاسازی و استخراج،

که در آن $d(C, C')$ فاصله اقلیدسی بین تصاویر پوشانه و حامل $\theta_{d2}, \theta_s, \theta_i, \theta_h, \theta_{d1}$ مولفه‌های شبکه و می‌شود تعریف می‌کنند.

که در آن $d(C, C')$ فاصله اقلیدسی بین تصاویر پوشانه و حامل $\theta_{d2}, \theta_s, \theta_i, \theta_h, \theta_{d1}$ مولفه‌های شبکه و



(شکل - ۱۱): مدل HCGAN [69]

Figure 11: The HCGAN model [69]

روش‌های نهان‌نگاری با حداقل خطای طبقه‌بندی ارزیابی می‌شود:

$$P_E = \min\left(\frac{P_{FA} + P_{MD}}{2}\right) \quad (75)$$

که در آن P_{FA} و P_{MD} به ترتیب احتمال هشدار کاذب و احتمال تشخیص نادرست هستند. نرخ تشخیص نهان‌کاوی با نرخ خطای میانگین P_E روی یک دسته تصادفی از مجموعه داده‌ها ارزیابی می‌شود و P_E بزرگ‌تر به معنای امنیت بیشتر روش نهان‌نگاری است.

۸-۱-۱- نهان‌نگاری جمع‌شونده

جدول ۱ میانگین احتمال کل خطای P_E و انحراف استاندارد آن را برای روش‌های نهان‌نگاری تطبیقی جمع‌شونده معرفی شده در بخش ۵ نشان می‌دهد. بررسی ادبیات فعلی نهان‌کاوی در حوزه مکان نشان می‌دهد که روش‌های اخیر نهان‌نگاری مبتنی بر هزینه مانند HILL [33] و مبتنی بر مدل مانند MiPOD [31] هر دو به سطح مشابهی از قابلیت تشخیص تجربی دست می‌یابند. با این حال، عملکرد آن‌ها متفاوت است و هزینه‌ها به‌طور معمول به‌صورت اکتشافی از طریق بازخورد ارائه شده توسط نهان‌کاوی تجربی به دست می‌آیند؛ اما در روش مبتنی بر مدل، نرخ تغییرات با به کمینه‌رساندن واگرایی KL بین توزیع‌های پوشانه و حامل محاسبه می‌شود. از طرف دیگر، در روش‌های مبتنی بر هزینه، ضریب انحراف نهان‌کاوی بر اساس نرخ تغییر خطی است؛ درحالی‌که در روش‌های مبتنی بر مدل، ضریب انحراف بر اساس نرخ تغییر درجه دوم است.

۸- بحث و بررسی

این بخش به بررسی روش‌های معرفی شده می‌پردازد. آزمایش‌ها بر روی مجموعه داده BOSSbase ver1.01 [70] شامل ۱۰۰۰۰ تصویر خاکستری هشت‌بیتی با ابعاد 512×512 انجام گرفته است. در شکل (۱۲) چند نمونه از این تصاویر مشاهده می‌شود.



(شکل - ۱۲): نمونه تصاویر مجموعه داده

BOSSbase ver1.01 [70]

Figure 12: The samples of BOSSbase ver1.01 [70]

۸-۱- روش‌های نهان‌نگاری سنتی

همه روش‌های مورد آزمایش با تنظیمات پیش‌فرض و در محدوده نظری شبیه‌سازی شده‌اند. نهان‌کاوها با استفاده از طبقه‌بند^۱ EC و تمایز خطی فشر^۲ (FLD) برای دو مجموعه ویژگی SRM [54] و maxSRMd2 [71] که نسخه آگاه از کانال انتخاب مجموعه SRM است و ویژگی‌ها را با احتمالات جاسازی تخمینی وزن‌دهی می‌کند، آموزش می‌بینند. از این‌رو، عملکرد امنیتی

¹ Ensemble classifier

² Fisher Linear Discriminant

³ Selection channel aware

Table 1: Detection error P_E of methods based on additive distortion [31]

Payload (bpp)						Method	Steganalyzer
0.5	0.4	0.3	0.2	0.1	0.05		
.1683±.0023	.2060±.0022	.2553±.0028	.3210±.0038	.4026±.0028	.4572±.0026	WOW	SRM
.1640±.0024	.2037±.0032	.2571±.0016	.3199±.0027	.4024±.0019	.4533±.0026	S -UNIWARD	
.1450±.0010	.1796±.0014	.2255±.0015	.2871±.0016	.3716±.0013	.4255±.0016	HUGO -BD	
.2055±.0024	.2482±.0030	.2996±.0022	.3611±.0024	.4364±.0034	.4691±.0017	HILL	
.1833±.0028	.2210±.0022	.2698±.0018	.3300±.0036	.4065±.0043	.4513±.0021	MiPOD	
.1119±.0029	.1357±.0030	.1658±.0024	.2146±.0028	.2953±.0026	.3689±.0019	MVG	
.1306±.0021	.1543±.0036	.1886±.0036	.2339±.0041	.2997±.0023	.3539±.0024	WOW	maxSRMd2
.1551±.0019	.1908±.0025	.2360±.0022	.2886±.0025	.3660±.0040	.4180±.0025	S -UNIWARD	
.1326±.0007	.1635±.0014	.2020±.0015	.2431±.0018	.3130±.0025	.3652±.0023	HUGO -BD	
.1814±.0030	.2184±.0037	.2573±.0033	.3091±.0018	.3771±.0019	.4232±.0029	HILL	
.1678±.0038	.2038±.0039	.2481±.0027	.3030±.0019	.3747±.0014	.4300±.0028	MiPOD	
.0715±.0018	.0813±.0018	.0936±.0015	.1161±.0016	.1653±.0019	.2315±.0027	MVG	

$$\Lambda(x) = \sum_{i=1}^n \Lambda_i = \sum_{i=1}^n \log \left(\frac{q_{Y_i}(x_i)}{p(x_i)} \right) \underset{H_0}{\overset{H_1}{\geq}} \tau \quad (77)$$

در نتیجه

$$q = \frac{\sum_{i=1}^n (E_{H_1}[\Lambda_i] - E_{H_0}[\Lambda_i])}{\sqrt{\sum_{i=1}^n \text{Var}_{H_0}[\Lambda_i]}} = \frac{\sqrt{2} \sum_{i=1}^n c_i \pi_i \gamma_i}{\sqrt{\sum_{i=1}^n c_i \gamma_i^2}} \quad (78)$$

که q بیانگر ضریب انحراف است و هدف نهان‌کاو به حداکثر رساندن انحراف است درحالی‌که در نهان‌نگاری تلاش می‌شود انحراف به حداقل برسد. با در نظر گرفتن نهان‌نگاری به‌عنوان یک بازی مجموع صفر^۴ بین نهان‌نگار و نهان‌کاو، در حالت تعادل، فرستنده باید نرخ‌های تغییری π_i را انتخاب کند که قابلیت تشخیص آماری را به حداقل برساند که به‌طور مستقیم با ضریب انحراف مرتبط است. لازم به ذکر است، اجرای چنین آزمونی در عمل آسان نیست؛ زیرا اغلب تخمین دقیقی از توزیع‌های $p(x_i)$ و $q(x_i)$ در دسترس نیست یا ابعاد نمایش تصویر خیلی بزرگ است که این مسئله را از نظر محاسباتی دشوار یا غیرممکن می‌کند.

روش‌های نهان‌کاو به دو دسته آگاه و ناآگاه تقسیم می‌شوند. نهان‌کاو ناآگاه نسبت به اقدامات نهان‌نگار کاملاً بی‌اطلاع است و در نتیجه نرخ تغییر را برای همه پیکسل‌ها ثابت در نظر می‌گیرد $\gamma_i = \gamma$ درحالی‌که نهان‌کاو آگاه از نرخ تغییرات مورد استفاده نهان‌نگار π_i مطلع است و از آن برای هدف خود بهره می‌برد $\gamma_i = \pi_i$. نهان‌کاو آگاه بدترین سناریو را برای طراحی نهان‌نگاری نشان می‌دهد.

^۴ zero-sum game

- ضریب انحراف نهان‌کاو

فرض می‌شود که نهان‌کاو از نرخ تغییر γ $(\gamma_1, \dots, \gamma_n)$ استفاده می‌کند که ممکن است با $\pi = (\pi_1, \dots, \pi_n)$ منطبق باشد. هنگام تجزیه و تحلیل تصویر پوشانه، هدف نهان‌کاو بهینه‌سازی برخی از معیارهای آزمون فرضیه خود جهت تصمیم‌گیری بین دو فرضیه ساده زیر خواهد بود:

$\mathcal{H}_0$: شیء مورد بررسی یک پوشانه است.

$\mathcal{H}_1$: شیء مورد بررسی یک حامل است.

هدف نهان‌کاو شناسایی یک نگاشت $\delta: Z^n \rightarrow \{H_0, H_1\}$ ، با بهترین عملکرد ممکن است. با بهره‌گیری از معیار بهینه‌سازی نیمن-پیرسون، برای احتمال هشدار کاذب^۱ مشخص $\alpha_0 = \mathbb{P}(\delta(x) = \mathcal{H}_0 | \mathcal{H}_1)$ به دنبال آزمونی است که تابع زیر را بهینه کرده و احتمال تشخیص صحیح را به دست آورد.

$$\mathbb{P}(\delta(x) = \mathcal{H}_1 | \mathcal{H}_1) \quad (76)$$

لم نیمن-پیرسون^۲ آزمون نسبت درست‌نمایی^۳ را به‌عنوان قوی‌ترین آزمون که حداکثر توان عملکرد را برای احتمال هشدار کاذب به دست می‌آورد معرفی می‌کند. آزمون نسبت درست‌نمایی یکی از روش‌های آزمون فرض آماری است که بین درستی یک فرضیه و متمم آن تصمیم می‌گیرد. با فرض استقلال آماری پیکسل‌ها:

^۱ False-alarm probability^۲ Neyman-pearson lemma^۳ Likelihood Ratio Test (LRT)

برای نهان‌کاو ناآگاه، تعریف اعوجاج به صورت زیر ساده می‌شود:

$$q = \frac{\sqrt{2} \sum_{i=1}^n c_i \pi_i}{\sqrt{\sum_{i=1}^n c_i}} \quad (79)$$

با توجه به این که مخرج ثابت است می‌توان آن را نادیده گرفت که در آن ضریب انحراف با احتمالات جاسازی به صورت خطی در ارتباط است.

$$q \propto \sum_{i=1}^n c_i \pi_i \quad (80)$$

همان‌طور که مشاهده می‌شود این رابطه معادل با میانگین اعوجاج در رویکرد مبتنی بر هزینه است؛ بنابراین، در رویکرد مبتنی بر هزینه فرض بر ناآگاه بودن نهان‌کاو است.

از طرف دیگر، برای نهان‌کاو آگاه که از محتوای پوشانه و نرخ تغییرات اطلاع دارد، ضریب انحراف بر اساس مربع احتمالات جاسازی تعریف می‌شود:

$$q = \frac{\sqrt{2} \sum_{i=1}^n c_i \pi_i^2}{\sqrt{\sum_{i=1}^n c_i \pi_i^2}} = \sqrt{2 \sum_{i=1}^n c_i \pi_i^2} \quad (81)$$

$$q^2 \propto \sum_{i=1}^n c_i \pi_i^2 \quad (82)$$

که معادل با ضریب انحراف در رویکرد مبتنی بر مدل است؛ بنابراین در این رویکرد فرض می‌شود که نهان‌کاو از نرخ تغییرات استفاده شده توسط نهان‌نگار آگاه است.

طبق نتایج تجربی که در [۷۲] بیان شده است، روش‌های مبتنی بر مدل انطباق کمتری با محتوا داشته و حتی گاهی تغییرات با هزینه بالا را به ایجاد تغییرات کم‌هزینه ترجیح می‌دهند در نتیجه با احتمال بیشتری نسبت به روش‌های مبتنی بر هزینه، پیام را در مناطق هموار جاسازی می‌کنند، بنابراین قابلیت تشخیص را به‌ویژه در برابر نهان‌کاوهای ناآگاه افزایش می‌دهد. روش‌های مبتنی بر مدل محدودیت‌های مهمی دارند زیرا

دانش قابل توجهی از جمله هزینه‌های اعوجاج دقیق (حتی با وجود در دسترس نبودن پوشانه) و مدل پوشانه را برای نهان‌کاو در نظر می‌گیرند. اگر شرایطی وجود داشته باشد که نهان‌کاو این دانش را در اختیار داشته باشد، ممکن است تابع هدف رویکرد مبتنی بر هزینه (۵) نتیجه بهینه حاصل نکند؛ زیرا معیار ارزیابی، ویژگی توزیع‌های پوشانه و حامل است و نمی‌تواند از کمینه‌سازی اعوجاج در تصاویر منفرد حاصل شود. در این صورت، رویکرد مبتنی بر مدل مفید خواهد بود؛ اما با توجه به اینکه این دانش در عمل برای نهان‌کاو در دسترس نیست، بنابراین، برای کاربردهای عملی روش‌های نهان‌نگاری مبتنی بر هزینه ترجیح داده می‌شوند.

۸-۱-۲- همگام‌سازی تغییرات جاسازی

در جدول (۲) عملکرد روش‌های مبتنی بر همگام‌سازی تغییرات مانند DeJoin, Synch, CMD و AsymMF برای ظرفیت‌های مختلف با هم مقایسه می‌گردند. برای دستیابی به یک مقایسه منصفانه، HILL به‌عنوان اعوجاج اولیه برای همه روش‌ها استفاده شده است.

با توجه به این که نهان‌کاوها تنها می‌توانند نرخ تغییرات حاصل از یک روش جمع‌شونده را به دست آورند، بنابراین در روش‌های مبتنی بر همگام‌سازی، تغییرات انجام گرفته در زیرتصویر نخست که هزینه‌ها در آن توسط روش HILL محاسبه شده است، قابل‌شناسایی هستند و احتمال تغییر پیکسل‌ها در سایر زیرتصاویر به دلیل وابستگی به تغییرات زیرتصاویر دیگر، ممکن است در فرآیند جاسازی بسیار متفاوت باشد.

جدول ۲: خطای تشخیص P_E روش‌های مبتنی بر اعوجاج غیرجمع‌شونده با جاسازی نامتقارن [۳۹]

Table 2: Detection error P_E based on non-additive distortion with asymmetric embedding [39]

Payload (bpp)					Method	Steganalyzer
0.5	0.4	0.3	0.2	0.1		
0.2374	0.2846	0.3267	0.3825	0.4482	Synch	SRM
0.2496	0.2885	0.3358	0.3841	0.4508	DeJoin2	
0.2577	0.2963	0.3483	0.3945	0.4572	CMD	
0.2623	0.2987	0.3492	0.3875	0.4524	DeJoin4	
0.2847	0.3279	0.3646	0.4112	0.4685	AsymMF4	
0.2237	0.2546	0.2894	0.3313	0.3867	Synch	maxSRMd2
0.2258	0.2576	0.2957	0.3392	0.393	DeJoin2	
0.2354	0.2672	0.3048	0.3473	0.4012	CMD	
0.2381	0.2713	0.3073	0.3407	0.3943	DeJoin4	
0.2517	0.2844	0.3175	0.3581	0.4136	AsymMF4	



به MiPOD نشان می‌دهد؛ اما برای ظرفیت‌های متوسط و بزرگ عملکرد GMRF ضعیف‌تر است ($\alpha \geq 0.3\text{bpp}$). بهبود عملکرد GMRF نسبت به روش مستقل MiPOD عمدتاً در مناطق با بافت متوسط و کم دیده می‌شود. همگام‌سازی تغییرات به دو دلیل باعث بهبود امنیت نهان‌نگاری می‌شود. نخست، با توجه به اینکه اغلب پیش‌بینی‌کننده‌های مورد استفاده برای استخراج باقیمانده‌های نوفه در نهان‌کاوها، مانند مدل غنی فضایی (SRM)، از فیلترهایی با علائم متناوب استفاده می‌کنند؛ ازاین‌رو، با تغییرات در جهات یکسان کمتر دچار اختلال می‌شوند. دوم، از آنجایی که جهت‌های جاسازی به تغییرات دقیق جاسازی بستگی دارد، کانال انتخاب (نقشه احتمالات تغییر) در دسترس نهان‌کاو نیست که این امر باعث کاهش کارایی نهان‌کاوهای آگاه از کانال مانند مجموعه ویژگی maxSRMd2 می‌شود.

۸-۱-۳- هموارسازی هزینه‌های جاسازی

طبق پژوهش‌های انجام‌شده در [۳۳]، هموارسازی هزینه‌های جاسازی با استفاده از یک فیلتر پایین‌گذر باعث بهبود محرمانگی روش‌های نهان‌نگاری می‌شود. طبق اصل بیشینه آنتروپی در تئوری اطلاعات، عملیات هموارسازی هزینه‌ها منجر به یکنواخت‌تر شدن تغییرات در مناطق محلی و در نتیجه افزایش آنتروپی در مناطق با بافت غنی می‌شود. این امر به دلیل کاهش تغییر در مشخصات آماری پوشانه، محرمانگی و امنیت را ارتقا می‌بخشد.

ازاین‌رو، با توجه به این که پیکسل‌ها در روش‌های Synch و DeJoin2 به دو زیرمجموعه و در CMD، DeJoin4 و AsymMF4 به چهار زیرمجموعه تقسیم می‌شوند، بنابراین نهان‌کاو می‌تواند احتمال تغییر نیمی از پیکسل‌ها را برای Synch و DeJoin2 و تنها یک‌چهارم پیکسل‌ها را برای CMD و DeJoin4 تخمین بزند. ازاین‌رو، روش‌های DeJoin4 و CMD عملکرد مشابهی داشته و نتایج بهتری نسبت به دیگر روش‌ها به دست می‌آورند. برای روش AsymMF4، به دلیل تعریف تابع اعوجاج غیر جمع‌شونده، احتمال تغییر برای کل تصویر غیر قابل تخمین است و این امر باعث بهبود نتایج این روش نسبت به بقیه روش‌ها گردیده است.

در راستای همگام‌سازی تغییرات جاسازی بر اساس تعاملات متقابل پیکسل‌ها، مدل GMRF [۳۷] بر اساس رویکرد مبتنی بر مدل معرفی شده است. در طراحی روش‌های نهان‌نگاری غیرجمع‌شونده پیشین، الگوریتم جاسازی ماهیت نامتقارن دارد که نرخ تغییر متفاوت برای پیکسل‌ها در تغییرات $+1/-1$ محاسبه می‌کند. درحالی‌که در GMRF جاسازی متقارن در نظر گرفته شده است. ازاین‌رو، GMRF می‌تواند به عنوان اعوجاج اولیه در روش‌های غیرجمع‌شونده مانند CMD و Synch مورد استفاده قرار گیرد که باعث بهبود عملکرد آن خواهد شد.

طبق نتایج بیان‌شده در جدول (۳) مشخص می‌شود که مدل GMRF می‌تواند توزیع آماری تصاویر حامل را پس از جاسازی تا حد قابل قبولی حفظ کند. برای مجموعه ویژگی maxSRMd2، GMRF ظرفیت‌های کم ($\alpha \leq 0.2\text{bpp}$) عملکرد بهتری نسبت

(جدول-۳): خطای تشخیص P_E روش‌های MiPOD و GMRF و بررسی تأثیر همگام‌سازی تغییرات

در روش‌های مبتنی بر مدل [۳۷]

Table 3: Detection error P_E of MiPOD and GMRF methods and evaluation of the effect of synchronizing the embedding changes in model based methods [37]

Payload (bpp)						Method	Steganalyzer
0.5	0.4	0.3	0.2	0.1	0.05		
0.1821±0.0023	0.2218±0.0019	0.2723±0.0041	0.3274±0.0026	0.4051±0.0041	0.4511±0.0030	MiPOD	SRM
0.1925±0.0028	0.2350±0.0035	0.2860±0.0043	0.3428±0.0038	0.4143±0.0036	0.4558±0.0024	GMRF	
0.1683±0.0037	0.2053±0.0028	0.2498±0.0034	0.3037±0.0028	0.3772±0.0028	0.4294±0.0037	MiPOD	maxSRMd2
0.1595±0.0033	0.1949±0.0041	0.2467±0.0035	0.3078±0.0036	0.3830±0.0020	0.4361±0.0031	GMRF	

(جدول-۴): خطای تشخیص P_E روش‌های HILL, MiPOD, GMRF و SymMF (بررسی تاثیر هموارسازی هزینه‌ها) [۳۹]

Table 4: Detection error P_E of HILL, MiPOD, GMRF and SymMF methods (Investigating the effect of costs smoothing) [39]

Payload (bpp)						Method	Steganalyzer
0.5	0.4	0.3	0.2	0.1	0.05		
0.2018±0.0027	0.2450±0.0025	0.2975±0.0055	0.3582±0.0026	0.4330±0.0029	0.4704±0.0024	HiLL	SRM
0.1910±0.0020	0.2362±0.0039	0.2811±0.0048	0.3417±0.0031	0.4139±0.0031	0.4549±0.0034	MiPOD	
0.2005±0.0023	0.2444±0.0033	0.2948±0.0020	0.3530±0.0031	0.4210±0.0042	0.4581±0.0027	GMRF	
0.2178±0.0029	0.2683±0.0025	0.3136±0.0042	0.3746±0.0031	0.4412±0.0035	0.4623±0.0041	SymMF	
0.1789±0.0033	0.2169±0.0018	0.2590±0.0029	0.3094±0.0029	0.3732±0.0045	0.4237±0.0019	HiLL	maxSRMd2
0.1846±0.0029	0.2212±0.0023	0.2684±0.0026	0.3231±0.0036	0.3902±0.0028	0.4416±0.0030	MiPOD	
0.1780±0.0024	0.2201±0.0038	0.2684±0.0035	0.3240±0.0020	0.3958±0.0032	0.4445±0.0020	GMRF	
0.2052±0.0024	0.2349±0.0043	0.2714±0.0018	0.3207±0.0036	0.3881±0.0029	0.4722±0.0027	SymMF	

و امنیت را افزایش می‌دهد. به‌طور خلاصه، عملیات هموارسازی مقاومت را، به‌ویژه در برابر نهان‌کاوهای آگاه از کانال مانند maxSRMd2، افزایش می‌دهد.

۴-۱-۸- استفاده از اطلاعات جانبی

نهان‌نگاری با اطلاعات جانبی عموماً تغییر مؤلفه‌هایی از پوشانه که خطای گرد کردن آن‌ها نزدیک به $\pm 1/2$ است را ترجیح می‌دهد زیرا چنین عناصری به اغتشاشات کوچک حساس‌تر هستند. به‌عنوان مثال، یک مؤلفه پوشانه با مقدار غیر گرد $2/57$ که به ۳ گرد می‌شود، در حین جاسازی با یک هزینه کم به ۲ اصلاح می‌شود درحالی‌که تغییر مؤلفه پوشانه با مقدار ۳ به ۴ هزینه نسبتاً بزرگ‌تری را به همراه دارد.

در جدول (۴) عملکرد روش‌های HILL و SymMF که از هموارسازی برای تعیین هزینه‌ها بهره گرفته‌اند با روش‌های GMRF و MiPOD که با یک فیلتر پایین‌گذر ترکیب شده‌اند، مورد مقایسه قرار می‌گیرد. همان‌طور که از نتایج مشخص است، برای SRM، GMRF با فیلتر پایین‌گذر نسبت به GMRF بدون فیلتر و MiPOD عملکرد بهتری دارد؛ اما عملکرد آن کمی پایین‌تر از HILL است. با افزایش ظرفیت، شکاف عملکرد بین GMRF و HILL کمتر می‌شود. در روش SymMF، تعاملات بین پیکسل‌ها با کمک MRF مدل شده است که امکان تعریف ارتباطات مختلف بین متغیرها را فراهم می‌سازد؛ سپس، با بهره‌گیری از روش میدان متوسط، در قالب فیلتر هموارکننده تطبیقی، هزینه‌های تغییر یک روش پایه مانند HILL به‌روزرسانی می‌شود. طبق نتایج نشان داده شده در جدول ۴، این امر باعث کاهش تغییرات آماری حاصل از جاسازی پیام در پوشانه گردیده

جدول ۵: بررسی تاثیر ترکیب دو رویکرد همگام‌سازی تغییرات ناشی از جاسازی و استفاده از اطلاعات جانبی [۴۳]

Table 5: Evaluation of the effect of synchronizing the embedding changes and using the sided information [43]

Payload (bpp)					Method	Steganalyzer
0.5	0.4	0.3	0.2	0.1		
0.2115±0.0019	0.2552±0.0026	0.3063±0.0031	0.3678±0.0036	0.4356±0.0022	HILL	SRM
0.2631±0.0027	0.3015±0.0034	0.3498±0.0023	0.4001±0.0032	0.4548±0.0021	CMD-HILL	
0.2833±0.0018	0.3338±0.0034	0.3893±0.0019	0.4351±0.0025	0.4747±0.0012	SI-HILL	
0.3263±0.0026	0.3673±0.0027	0.4075±0.0024	0.4426±0.0022	0.4742±0.0012	nmSI-HILL	
0.2061±0.0013	0.2640±0.0033	0.2948±0.0033	0.3510±0.0021	0.4206±0.0021	MiPOD	



0.2590±0.0025	0.2985±0.0025	0.3439±0.0030	0.3912±0.0022	0.4419±0.0026	CMD-MIPOD	
0.2725±0.0033	0.3181±0.0029	0.3672±0.0028	0.4139±0.0031	0.4578±0.0023	SI-MIPOD	
0.3109±0.0025	0.3543±0.0033	0.3897±0.0022	0.4242±0.0015	0.4611±0.0017	nmSI-MIPOD	
0.1887±0.0023	0.2264±0.0015	0.2703±0.0030	0.3188±0.0030	0.3863±0.0014	HILL	maxSRMd2
0.2413±0.0025	0.2681±0.0016	0.3142±0.0030	0.3568±0.0030	0.4125±0.0030	CMD-HILL	
0.2669±0.0035	0.3191±0.0022	0.3740±0.0020	0.4336±0.0019	0.4832±0.0031	SI-HILL	
0.3096±0.0026	0.3509±0.0029	0.3983±0.0032	0.4442±0.0033	0.4818±0.0020	nmSI-HILL	
0.1926±0.0016	0.2313±0.0017	0.2750±0.0026	0.3331±0.0019	0.4002±0.0018	MiPOD	
0.2425±0.0023	0.2793±0.0020	0.3233±0.0030	0.3669±0.0016	0.4249±0.0023	CMD-MIPOD	
0.2680±0.0030	0.3192±0.0023	0.3743±0.0017	0.4344±0.0024	0.4805±0.0021	SI-MIPOD	
0.3048±0.0026	0.3527±0.0026	0.3952±0.0023	0.4463±0.0034	0.4806±0.0022	nmSI-MIPOD	

مبتنی بر هزینه یک مدل آماری و روشی برای تخمین مدل و اطلاعات فیشر آن وجود دارد.

در [۷۳] الگوریتم HILL با یک مدل گاوسی پیاده‌سازی می‌شود که اساساً نسخه‌ای از MiPOD با تخمین گر واریانس متفاوت است. در این روش، هزینه‌ها به‌عنوان برآوردهای متقابل انحراف استاندارد محلی تفسیر می‌شوند. روش HILL هزینه‌ها را به‌صورت اکتشافی با استفاده از یک سری فیلتر محاسبه می‌کند. با نادیده گرفتن فیلتر پایین‌گذر دوم برای سادگی، هزینه‌ها را می‌توان به‌عنوان میانگین متقابل قدر مطلق باقیمانده $\rho_i \approx 1/E[|R_i|]$ یا میانگین انحراف مطلق^۱ و با فرض صفر بودن باقیمانده KB در نظر گرفت. برای طیف وسیعی از توزیع‌ها، میانگین قدر مطلق با فرض ثابت بودن پارامترهای دیگر، با انحراف استاندارد متناسب است $E[X] \propto \sigma_i$ ؛ بنابراین، هزینه متقابل به‌صورت $\frac{1}{\rho_i} \approx E[|R_i|] \propto \sigma_i$ بیان می‌شود.

با یک مدل باقیمانده گاوسی محلی، نسخه متفاوتی از MiPOD با تخمین گر واریانس مبتنی بر HILL $(\sigma_i^2 = \frac{\pi}{2\rho_i^2})$. این نسخه واریانس محلی را از هزینه‌های HILL به دست می‌آورد که منجر به بهبود امنیت، به‌ویژه برای نهان‌کاو آگاه می‌شود؛ درحالی‌که برای ظرفیت‌های کمتر بهبود بیشتری دیده می‌شود (جدول ۶). همچنین، استفاده از هزینه‌های HILL برای تخمین واریانس از نظر حفظ توزیع کلی باقیمانده نسبت به تخمین گر واریانس اصلی در MiPOD بهتر است.

۸-۲- نهان‌نگاری مبتنی بر یادگیری تقابلی

در این بخش به بررسی روش‌های نهان‌نگاری مبتنی بر یادگیری عمیق پرداخته می‌شود. امنیت و محرمانگی به معنای غیرقابل تشخیص بودن تصاویر نهان‌نگاری در

گفتنی است اطلاعات جانبی به‌جز خطاهای گرد کردن می‌تواند از روش‌های دیگری نیز به دست آید. به‌عنوان مثال، هنگامی‌که فرستنده به چندین نوردهی از یک صحنه دسترسی داشته باشد می‌تواند از آن‌ها برای تخمین جهت تغییرات جاسازی برای مؤلفه‌های پوشانه که بیشتر مستعد نویزهای کوچک هستند استفاده نماید و بنابراین بهتر می‌تواند تغییرات جاسازی را به‌عنوان نویز اکتسابی تقلید کند.

در جدول ۵ تاثیر استفاده از اطلاعات جانبی (SI)، خوشه‌بندی تغییرات (CMD) و ترکیب این دو روش (nmSI) بر روی روش‌های پایه مبتنی بر هزینه مانند HILL و مبتنی بر مدل مانند MiPOD مورد بررسی قرار گرفته است. همان‌طور که از نتایج مشخص است، استفاده از اطلاعات جانبی به‌دلیل این‌که این اطلاعات فقط در دسترس فرستنده است، می‌تواند امنیت را تا حد زیادی بهبود بخشد. همچنین، با لحاظ نمودن تعاملات بین پیکسل‌ها و خوشه‌بندی تغییرات نتایج بهتری نسبت به روش‌هایی که فقط از خوشه‌بندی تغییرات و یا اطلاعات جانبی استفاده کرده‌اند، حاصل می‌شود.

۸-۱-۵- تبدیل روش‌های مبتنی بر هزینه به

مدل

توسعه توابع اعوجاج کمتر اکتشافی یک حوزه مهم برای تحقیقات نهان‌نگاری است. در این راستا، در [۶۸،۶۹] امکان تبدیل روش‌های مبتنی بر هزینه به مدل بررسی شده و نتایج نشان می‌دهد که به ازای یک ظرفیت مشخص، نرخ تغییر ناشی از هزینه‌ها با نرخ تغییر حاصل از برخی مدل‌ها منطبق است. این معیار نسبتاً ساده گاهی پیشرفت قابل‌توجهی در امنیت به‌خصوص برای نهان‌کاو آگاه به همراه دارد. به بیان دیگر، برای روش

¹ Mean Absolute Deviation

قرار می‌گیرند. لازم به ذکر است، به دلیل محدودیت‌های محاسبات GPU و زمان، در اغلب راهکارهای مبتنی بر یادگیری عمیق، اندازه تصاویر از 512×512 به 256×256 کاهش یافته و آزمایش‌های این بخش بر روی تصاویر با ابعاد جدید انجام می‌شود.

برابر ابزارهای نهان‌کاوی است. روش‌های نهان‌کاوی برای ارزیابی خصوصیات آماری تصاویر، ویژگی‌های مقاوم به تغییرات جاسازی را استخراج می‌کنند. با توجه به اهمیت انتخاب ویژگی در رویکرد نهان‌کاوی، محققان به‌تازگی از شبکه‌های عصبی کانولوشنی (CNN) جهت استخراج ویژگی‌های مناسب سود می‌برند. از این‌رو، اغلب روش‌های نهان‌نگاری مبتنی بر یادگیری عمیق توسط این نهان‌کاوها مورد ارزیابی

جدول ۶: خطای تشخیص P_E روش MiPOD با واریانس بدست آمده از روش HILL و روش MiPOD اصلی [۷۳]
Table 6: Detection error P_E for MiPOD with HILL-based variance estimator and the original MiPOD estimator [73]

Payload (bpp)				Steganalyzer	Method
0.4	0.3	0.2	0.1		
0.2213	0.2678	0.3206	0.3937	maxSRMd2	HILL -inspired
0.2142	0.2552	0.3101	0.3800	maxSRMd2	MiPOD
0.1545	0.1870	0.2470	0.3390	SRNet	HILL -inspired
0.1420	0.1826	0.2354	0.3575	SCA -SRNet	
0.1146	0.1553	0.2222	0.3213	SRNet	MiPOD
0.1106	0.1384	0.1961	0.2952	SCA -SRNet	

[58] EMB از یک روش جستجو برای یافتن بهترین الگوی تغییر استفاده می‌کند به‌طوری‌که الگوی به‌دست‌آمده بهترین عملکرد امنیتی را در برابر نهان‌کاو داراست. در این روش، فرآیند تخصیص هزینه با تنظیم نامتقارن بخشی از هزینه‌های جاسازی با توجه به گرادین حاصل از نهان‌کاو هدف انجام می‌گیرد. برای جلوگیری از تغییرات غیرضروری، سعی می‌شود تعداد عناصر با هزینه‌های قابل تنظیم به حداقل برسد. روش [60]، برای گسترش تنوع حامل، ابتدا چند حامل کاندید ایجاد و از بین آن‌ها حامل بهینه را انتخاب می‌کند. در روش [59] AEN، برخلاف روش‌های تقابلی دیگر، نه‌تنها از علامت سیگنال گرادین بلکه از مقدار سیگنال گرادین نیز برای اصلاح هزینه‌های جاسازی استفاده می‌شود که باعث بهبود عملکرد نهان‌نگاری می‌شود جدول (۸). تحقیقات نشان می‌دهد که روش‌های مبتنی بر نمونه‌های تقابلی می‌توانند برای بهبود روش‌های نهان‌نگاری مبتنی بر کمینه‌سازی اعوجاج مانند HILL و S-UNIWARD مؤثر باشند.

۸-۲-۱- تخمین ماتریس ویرایش

در جدول ۷ روش‌های UT-GAN، ASDL-GAN و CF-GAN که ماتریس احتمال ویرایش را بر اساس شبکه GAN محاسبه می‌کنند با روش‌های سنتی مقایسه می‌شود. نتایج تجربی نشان می‌دهد که روش UT-GAN عملکرد امنیتی را به‌طور چشمگیری در برابر روش‌های سنتی و همچنین روش ASDL-GAN بهبود می‌بخشد. دلیل آن این است که شبکه مولد در برابر نهان‌کاو مبتنی بر CNN آموزش دیده است. از طرف دیگر، روش CF-GAN با اضافه کردن کانال بازخورد متقاطع به شبکه باعث انتقال بهتر اطلاعات بین لایه‌های شبکه شده و این امر عملکرد روش را بهبود می‌بخشد.

۸-۲-۲- روش‌های مبتنی بر نمونه‌های تقابلی

در روش‌های مبتنی بر نمونه‌های تقابلی، هزینه تغییرات با توجه به سیگنال گرادین منتشرشده توسط نهان‌کاو هدف به‌روز می‌شود. در این راستا، جهت تغییر با معکوس جهت گرادین هم‌راستا می‌شود. روش-ADV

Table 8: Evaluation of the effect of using adversarial examples for both HILL and S-UNIWARD methods [52]

Payload (bpp)					Method	Steganalyzer
0.4	0.3	0.2	0.1	0.05		
0.195	0.25	0.32	0.42	0.45	S-UNIWARD	SRM
0.2	0.251	0.325	0.42	0.451	AEN-SUNIWARD	
0.225	0.28	0.355	0.43	0.47	HILL	
0.226	0.29	0.36	0.43	0.471	AEN-HILL	

بر چندین ویژگی تأثیر می‌گذارد، از اهمیت ویژه‌ای برخوردار است و یک مسیر تحقیق امیدبخش برای بهبود امنیت محسوب می‌شود.

۲- روش‌های نهان‌نگاری مبتنی بر هزینه که از اطلاعات جانبی در فرستنده استفاده نمی‌کنند، مانند HILL، تقریباً سطح امنیت تجربی یکسانی با روش‌های مبتنی بر مدل مانند MiPOD نشان می‌دهند. باین‌حال، آن‌ها بسیار متفاوت هستند، روش‌های مبتنی بر هزینه تابع هدف را به حداقل می‌رسانند که در نرخ تغییر خطی است درحالی‌که روش‌های مبتنی بر مدل انحراف را به حداقل می‌رساند که در نرخ تغییر درجه دوم است. از طرف دیگر، نهان‌نگاری مبتنی بر هزینه منجر به جاسازی «بیش از حد تطبیقی» می‌شود و به دشمن اجازه می‌دهد تا دقت تشخیص خود را با استفاده از اطلاعات کانال انتخاب بهبود بخشد. ازاین‌رو، توسعه توابع اعوجاج کمتر اکتشافی یک حوزه مهم برای تحقیقات آینده در نهان‌نگاری خواهد بود.

۳- بیشتر روش‌های مبتنی بر اعوجاج، احتمال تغییر پیکسل‌های پوشانه را مستقل فرض می‌کنند در حالی‌که در نظر گرفتن اثر متقابل بین تغییرات و همبستگی بین احتمالات تغییر در یک همسایگی محلی، از تولید الگوهای قابل‌شناسایی توسط نهان‌کاو جلوگیری کرده و محرمانگی را بهبود می‌بخشد؛ اما اکثر روش‌هایی که در این راستا معرفی شدند در دسته روش‌های جاسازی نامتقارن قرار می‌گیرند که به دلیل پیروی از آنتروپی شرطی نمی‌توانند حداکثر پیام را جاسازی کنند. ازاین‌رو، ارائه روشی که با وجود بهره‌گیری از تعاملات بین تغییرات مجاور، جاسازی را به صورت متقارن انجام دهد می‌تواند برای پژوهش‌های آینده مورد توجه قرار گیرد.

۸-۲-۳- روش‌های مبتنی بر بازی سه عامله

روش‌هایی که تاکنون بحث شد همه بر طراحی تابع اعوجاج متمرکز بوده و جاسازی پیام را به یک روش کدگذاری مؤثر مانند STC می‌سپارند که به‌طورمعمول از کم‌ارزش‌ترین بیت (LSB) برای جاسازی استفاده می‌کند. با توجه به اهمیت قوانین جاسازی برای امنیت نهان‌نگاری، به تازگی چارچوب‌های یکپارچه مبتنی بر بازی رقابتی بر اساس یادگیری عمیق مورد توجه قرار گرفته‌اند. با این‌که روش‌های مبتنی بر بازی سه عامله هنوز از لحاظ محرمانگی از روش‌های موجود نهان‌نگاری پیشی نگرفته‌اند اما این رویکرد پتانسیل پیشرفت داشته و راه را برای تحقیقات بیشتر هموار می‌کند.

۹- مسیرهای آینده

۱- نتایج نظری قدرتمندتر لزوماً منجر به نهان‌نگاری امن قابل اجرا نمی‌شود. اگرچه نتایج نظری کلی هستند اما محدودیت‌های مهمی دارند. آن‌ها دانش قابل‌توجهی از جمله هزینه‌های اعوجاج دقیق و مدل پوشانه را حتی باوجود در اختیار نداشتن پوشانه، برای آشکارساز در نظر می‌گیرند درحالی‌که یافتن هزینه‌های واقعی نیازمند دانستن پارامترهای دقیق مدل پوشانه است. محدودیت دیگر نیاز به یکسان بودن فضای تشخیص با فضایی است که در آن اعوجاج محاسبه می‌شود، بهینه بودن آزمون نسبت احتمال نشان می‌دهد که این دو در نهایت باید همگرا شوند؛ بنابراین به دلیل عدم اطلاع از هزینه‌های واقعی، نتایج نظری در خلأ درست هستند. باین‌حال، ممکن است برآوردهای خوبی از هزینه‌ها که می‌تواند از یادگیری تجربی از منبع پوشانه حاصل شود، برای یک تعادل تقریباً کافی باشد که می‌تواند در تحقیقات آینده موردتوجه قرار گیرد. همچنین انطباق نتایج نظری با موقعیت عملی که در آن یک تغییر جاسازی

(جدول - ۹): مرور روش‌های مورد مطالعه
Table 9: Review the steganographic methods

نمونه	مزایا و معایب	معرفی	رویکرد		نهان‌نگاری تطبیقی (تخمین بستر چسبازی پنهان)
			مبتنی بر مدل	مبتنی بر هزینه	
MVG MiPOD	نهان‌نگاری با این روش بسیار آسان خواهد بود؛ اما یافتن مدل مناسب دشوار است.	هدف این رویکرد حفظ مدل آماری پوشش است. در این رویکرد که بر اساس مشخصات آماری ارزیابی می‌گردد باید خصوصیات آماری پوشش و حاصل تا حد امکان یکسان باشد. و افروچ حاصل از چسبازی به حداقل برسد.	افروچ جمع‌شونده		با توجه به رابطه رفتاری نهان‌نگاری و نهان‌کاری، روش‌های نهان‌نگاری به‌عنوان یک حمله تقلبی سعی می‌کنند تا شبکه نهان‌کار را گمراه کنند تا همه تصاویر ورودی را به‌عنوان تصاویر پوشش تشخیص دهد. از این‌رو می‌توان از یادگیری تقلبی برای چسبازی مفهومی‌تر پیام بهره گرفت تا مشخصات آماری پوشش تا حد ممکن حفظ شود.
HUGO WOW UNWARD HILL	این رویکرد نسبت به رویکرد مبتنی بر مدل متعادلتر بوده و امکان توسعه روش‌های نهان‌نگاری را فراهم می‌کند. از این‌رو، از سطح امنیتی بهتری برخوردار است و رویکردی عملی‌تر است.	هدف این رویکرد به حداقل رساندن افروچ کل است. افروچ در این رویکرد مجموع هزینه‌های احتمالی مولف‌های ویرایش شده پوشش است.	افروچ غیرجمع‌شونده		مسئله پنهان‌سازی (کمینه سازی افروچ) نهان‌نگاری به عنوان یک مسئله پنهان‌سازی فرموله می‌شود که هدف آن به حداقل رساندن افروچ کل برای یک ظرفیت چسبازی مشخص است. افروچ معیاری است که توسط یک نهان‌کار برای مدل‌سازی اثر چسبازی پیام استفاده می‌شود و این روش‌ها سعی می‌کنند پیام را به‌گونه‌ای در پوشش مخفی کنند که مشخصات آماری پوشش تا حد ممکن حفظ گردد.
CMD Synch Dejoin GMRF SymMF AsymMF	روش‌های غیرجمع‌شونده با لحاظ نمودن اثر متقابل تغییرات محلی، از تولید الگوهای قابل‌شناسی توسط نهان‌کار جلوگیری کرده و محرمانگی را بهبود می‌بخشند.	روش‌های مبتنی بر افروچ جمع‌شونده، احتمالات تغییر پیکسل‌های پوشش را مستقل فرض می‌کنند؛ اما در واقعیت، تغییرات پیکسل‌های مجاور بر یکدیگر مؤثر می‌باشند. از این‌رو، لحاظ نکردن این وابستگی در عمل منجر به کاهش بهینگی می‌شود.	محاسبه ماتریس احتمال ویرایش		
ASDL-GAN UT-GAN CF-GAN	به دلیل رابطه رفتاری نهان‌نگاری و نهان‌کاری، استفاده از یک نهان‌کار به عنوان تمایزدهنده در GAN باعث یادگیری نقشه احتمال با قابلیت شناسایی کمتر و محرمانگی بالاتر می‌گردد.	با بهره‌گیری از GAN، مکان‌های مناسب برای چسبازی با کمک یادگیری نقشه احتمال شناسایی می‌شوند.	نهان‌نگاری با استفاده از نمونه‌های تقلبی		
ADV-EMB AEN	در این رویکرد، هزینه‌های تغییر بر اساس گردانیدن شبکه نهان‌کار اصلاح می‌گردند که باعث بهبود محرمانگی می‌گردد.	الطلب از مفهوم شبیه‌سازی بازی تقلبی با دو عامل فرستنده (نهان‌کار) و مهاجم (نهان‌کار) استفاده می‌کند تا هزینه‌های تغییر به‌دست‌آمده از روش‌های سنتی موجود را بهبود بخشد.	نهان‌نگاری شبیه‌سازی بازی سه‌عامله		
Yedroudi HCGAN	با توجه به اهمیت قوانین چسبازی برای امنیت نهان‌نگاری، از چارچوب یکپارچه مبتنی بر یادگیری تقلبی استفاده می‌کند. با این‌که روش‌های مبتنی بر بازی سه‌عامله هنوز از لحاظ محرمانگی از روش‌های موجود نهان‌نگاری پیشی نگرفته‌اند، اما این رویکرد با تسهیل پیشرفت دانش و راه را برای تحقیقات بیشتر هموار می‌کند.	با توجه به تأثیر فرایند چسبازی بر کیفیت تصویر، چارچوب نهان‌نگاری یکپارچه با الهام از مدل‌های یادگیری عمیق ارائه می‌شود که شامل یک مدل رمزنگار-رمزگشا برای یادگیری همزمان چسبازی و استخراج پیام‌های محرمانه است. این چارچوب در شبیه‌سازی بازی تقلبی از سه عامل استفاده می‌کند: فرستنده (شبکه نهان‌کار)، گیرنده (شبکه استخراج‌کننده) و مهاجم (شبکه نهان‌کار). در این بازی رفتاری، آموزش بین شبکه‌های نهان‌کار، استخراج‌کننده با هدف همگرایی با زبانی پیام از یک‌طرف و شبکه‌های نهان‌کار/نهان‌کار برای ایجاد یک تصویر حامل امن و غیرقابل‌شناسایی از طرف دیگر به‌صورت متناوب انجام می‌گیرد.	شبکه عصبی (یادگیری تقلبی)		

۴- همچنین روش‌های آماری می‌توانند در توسعه توابع اعوجاج برای حوزه‌هایی مانند صوت و ویدئو که کمتر مورد بررسی قرار گرفته‌اند، استفاده شوند.

۵- با توجه به پیشرفت شبکه‌های عمیق و حساسیت این شبکه‌ها به اغتشاشات کوچک تقابلی، می‌توان از یادگیری تقابلی برای جاسازی پیام بهره گرفت به‌طوری‌که مشخصات آماری پوشانه تا حد ممکن حفظ شود. اگرچه روش‌های مبتنی بر یادگیری عمیق تاکنون از لحاظ قابلیت تشخیص و محرمانگی در مقابل نهان‌کاوها، از روش‌های سنتی پیشی نگرفته‌اند، اما به دلیل وجود پارامترهای تصادفی نتایج قابل قبولی حاصل کرده‌اند. با توجه به مقداره‌ی اولیه تصادفی پارامترها در آموزش مدل، نداشتن مدل از قبل آموزش‌دیده دقت نهان‌کاوی را کاهش و امنیت را بهبود می‌دهد، از این‌رو، این رویکرد پتانسیل پیشرفت را دارد. در پژوهش‌های آینده، می‌توان امکان همگام‌سازی بیشتر فرستنده و گیرنده و همچنین فرستنده و مهاجم را مورد بررسی قرار داد. همچنین می‌توان تعریف تابع زیان را توسعه داد که به شبکه‌ها کمک کند تا به یک راه‌حل بهتر همگرا شوند.

۶- هدف اصلی نهان‌نگاری انتقال امن و محرمانه پیام است، به‌گونه‌ای که بازبایی کامل پیام توسط گیرنده تضمین می‌شود؛ درحالی‌که به دلیل ماهیت نایقینی شبکه‌های عصبی، استخراج دقیق پیام همراه با حفظ کیفیت و محرمانگی تصویر حامل، یکی از چالش‌های روش‌های نهان‌نگاری یکپارچه مبتنی بر یادگیری عمیق است. برای کنار آمدن با این مشکل، می‌توان از مزایای کدهای تصحیح خطا در پیش‌پردازش پیام استفاده کرد. استفاده از کدهای تصحیح خطا به‌طور قابل توجهی دقت پیام را بهبود می‌بخشد.

۱۰- نتیجه‌گیری

در این مقاله به بررسی مفهوم و ویژگی‌های نهان‌نگاری تصویر پرداخته و سپس با تمرکز بر یافتن بستر مناسب برای جاسازی پیام، روش‌های سنتی مبتنی بر بهینه‌سازی و روش‌های مبتنی بر یادگیری عمیق معرفی شد جدول (۹). در این راستا، ابتدا راه‌کارهای سنتی برای محاسبه نقشه احتمال جاسازی بهینه با هدف به کمینه‌رساندن اعوجاج ناشی از جاسازی پیام مورد

بررسی قرار گرفت. برای کمینه‌سازی اعوجاج، دو رویکرد مبتنی بر مدل و مبتنی بر هزینه معرفی شد که در رویکرد نخست یک مدل آماری برای پوشانه تعریف و سپس تلاش می‌شود در فرایند جاسازی پیام این مدل حفظ شود، درحالی‌که در رویکرد مبتنی بر هزینه هدف به کمینه‌رساندن اعوجاج حاصل از مجموع هزینه‌های اکتشافی به‌دست‌آمده از پوشانه است؛ پس از آن، به بررسی نهان‌نگاری مبتنی بر یادگیری عمیق پرداخته شد. این رویکرد بر اساس یادگیری تقابلی تعریف می‌شود که از رابطه رقابتی بین نهان‌نگار و نهان‌کاو برای بهبود عملکرد جاسازی بهره می‌گیرد. سه راه‌کار متفاوت در این حوزه معرفی شد. راه‌کار تخمین نقشه احتمال جاسازی با استفاده از یادگیری تقابلی از شبکه GAN برای یادگیری یک نقشه احتمال بهینه بهره می‌گیرد که به‌دلیل استفاده از نتایج شبکه نهان‌کاو در یادگیری، باعث بهبود محرمانگی و امنیت می‌شود. در راه‌کاری دیگر، از مفهوم نمونه‌های تقابلی برای اصلاح هزینه‌های به‌دست آمده از روش‌های سنتی استفاده می‌شود که این نیز باعث بهبود عملکرد نهان‌نگار می‌شود. در راه‌کار سوم، یک چارچوب یکپارچه مبتنی بر بازی سه عامله (نهان‌نگار، نهان‌کاو و گیرنده) ارائه می‌شود که به تازگی مورد توجه پژوهش‌گران بسیاری قرار گرفته است.

در سناریوی واقعی فرض بر این است که روش‌های نهان‌کاوی به الگوریتم نهان‌نگاری مورد بررسی، معماری مدل، ابرپارامترها و مجموعه تصاویر دسترسی دارند؛ اما لزوماً به پارامترهای دقیق مدل دسترسی ندارند. با توجه به مقداره‌ی اولیه تصادفی پارامترها در آموزش مدل، نداشتن مدل از قبل آموزش‌دیده باعث کاهش دقت نهان‌کاوی می‌شود. این ویژگی محرمانگی روش‌های مبتنی بر یادگیری تقابلی را بهبود می‌بخشد. از این‌رو، در حال حاضر توجه پژوهش‌گران به استفاده از رویکرد تقابلی در نهان‌نگاری جلب شده است.

از سوی دیگر، در روش‌های نهان‌نگاری مبتنی بر یادگیری عمیق، مکان تغییرات جاسازی شده به‌طور مستقیم پیام مخفی را منعکس نمی‌کند؛ بلکه نیازمند شبکه استخراج‌کننده برای بازبایی پیام از توزیع تغییرات است که این امر باعث بهبود محرمانگی و امنیت روش‌های یکپارچه مبتنی بر بازی سه عامله نسبت به روش‌هایی که از الگوریتم‌های جاسازی موجود استفاده می‌کنند، می‌شود؛ از این‌رو، امید است که این رویکرد به مسیرهای پرباری برای پژوهش‌های بیشتر منجر شود.

- [16] Pourmohammadali, A, Pourmohiabadi, M, Nezamabadi, H, "Safe steganography based on matrix embedding to increase embedding rate and efficiency," *Journal of Machine Vision and Image Processing*, No. 4, pp. 17-28, 2017
- [۱۷] فاتح، منصور، رجب‌لو، سمیرا، علی پور، الهه. "مروری بر نهان‌نگاری تصویر مبتنی بر مخفی‌سازی در کم‌ارزش‌ترین بیت و دسته‌بندی پیکسل و ارائه روشی جدید در این حوزه". امنیت فضای تولید و تبادل اطلاعات (منادی)، ۵ (۲): ۶۳-۷۱، ۱۳۹۵.
- [18] Fateh, M., Rezvani, M. and Irani, Y., A new method of coding for steganography based on LSB matching revisited. *Security and Communication Networks*, pp.1-15, 2021.
- [19] Bhardwaj, R. and Sharma, V., Image steganography based on complemented message and inverted bit LSB substitution. *Procedia Computer Science*, 93, pp.832-838, 2016.
- [20] Sahu, A.K. and Swain, G., High fidelity based reversible data hiding using modified LSB matching and pixel difference. *Journal of King Saud University-Computer and Information Sciences*, 34(4), pp.1395-1409, 2022.
- [21] Noorazar, A, Nowruzi, Z, Mir, M, "Providing an improved method for image steganography based on linear code features," *Electronic and cyber defense*. No. 5, pp. 43-53, 2017
- [۲۲] نورآذر، علی، نوروزی، زین‌العابدین و میر، مهدی، "ارائه روشی بهبودیافته برای نهان‌نگاری تصویر مبتنی بر ویژگی‌های کد‌های خطی"، *پدافند الکترونیکی و سایبری*. شماره ۵، ص ۴۳-۵۳، ۱۳۹۶
- [۲۳] ثابتی، وجیهه و احمدی، سارا، "نهان‌نگاری تطبیقی تصاویر در مقدار اختلاف ضرایب کسینوس گسسته"، *مجله علمی-پژوهشی رایانش نرم و فناوری اطلاعات*. شماره ۹، ص ۵۵-۶۶، ۱۳۹۸
- [24] Sabeti, V, Ahmadi, S, "Adaptive steganography of images in the difference of discrete cosine coefficients," *Journal of Soft Computing and Information Technology*. No. 9, pp. 55-66, 2019
- [25] Filler, T. and Fridrich, J., "Gibbs construction in steganography," *IEEE Transactions on Information Forensics and Security*, vol.5, no.4, pp.705-720, 2010.
- [26] Li, B., Tan, S., Wang, M., and Huang, J., "Investigation on cost assignment in spatial image steganography," *IEEE Transactions on Information Forensics and Security*, vol.9, no.8, pp.1264-1277, 2014.
- [27] Pevný, T., Filler, T., and Bas, P., "Using highdimensional image models to perform highly undetectable steganography," in *International Workshop on Information Hiding*, pp.161-177, Springer, 2010.
- [28] Holub, V. and Fridrich, J., "Designing steganographic distortion using directional filters," in *2012 IEEE International workshop*
- [1] Fridrich, J. *Steganography in digital media: principles, algorithms, and applications*. Cambridge University Press, 2009.
- [2] Kipper, G. *Investigator's guide to steganography*. crc press, 2003.
- [3] Bacon, F. and Watts, G., "Of the advancement and proficience of learning, or, the partitions of sciences, ix bookes," 1983.
- [4] Britannica, E., "A dictionary of arts, sciences, and general literature," *Edinburgh: Adam and Charles Black*, 1875.
- [5] Simmons, G. J., "The prisoners' problem and the subliminal channel," in *Advances in Cryptology*, pp.51-67, Springer, 1984.
- [6] Ker, A. D., Bas, P., Böhme, R., Cogranne, R., Craver, S., Filler, T., Fridrich, J., and Pevný, T., "Moving steganography and steganalysis from the laboratory into the real world," in *Proceedings of the first ACM workshop on Information hiding and multimedia security*, pp.45-58, 2013.
- [7] Holub, V. and others,. *Content Adaptive Steganography: Design and Detection*. Citeseer, 2014.
- [8] Fridrich, J., Soukal, D., "Matrix embedding for large payloads", *Information Forensics and Security*, IEEE Transactions on, Vol. 1, pp. 390-395, 2006.
- [9] Gao, Y., Li, X., Zeng, T., Yang, B., "Improving embedding efficiency via matrix embedding: a case study", in *Image Processing (ICIP), 16th IEEE International Conference on*, pp. 109-112, 2009.
- [10] Zhang, X., Zhang, W., Wang, S., "Efficient double-layered steganographic embedding", *Electronics letters*, Vol. 43, pp. 482-483, 2007.
- [۱۱] مهدوی، مجتبی، سماوی، شادرخ و خدامی، الهه، "نهان‌نگاری وفقی بر اساس پیچیدگی نسبی پیکسلها در تصاویر دوسطح"، *مجله مهندسی برق و الکترونیک ایران*. شماره ۶، ص ۳۷-۴۹، ۱۳۸۸
- [12] Mahdavi M, Samavi S, Khodami E. Steganography in Halftone Images based on Relative Complexity of Pixels. *Journal of Iranian Association of Electrical and Electronics Engineers*, 6 (1) :37-49, 2009.
- [۱۳] سلیمانی، سید رحمان، حق‌بین، سارا و نیازی، مسعود، "یک روش جدید نهان‌نگاری تطبیقی با ظرفیت مقیاس‌پذیر و کیفیت بصری بالا"، *فصلنامه علمی علوم و فناوریهای پدافند نوین*. شماره ۴، ص ۱-۱۴، ۱۳۹۲
- [14] Soleimani, R, Haghbin, S, Niazi, M, "A New Method of Comparative Steganography with Scalable Capacity and High Visual Quality," *Quarterly Journal of Modern Defense Science and Technology*. No. 4, pp. 1-14, 2013
- [۱۵] پورمحمدعلی، علیرضا، پورمحمی آبادی، مریم و نظام آبادی، حسین، "نهان‌نگاری ایمن مبتنی بر جاسازی ماتریسی جهت افزایش نرخ و بازده جاسازی"، *مجله ماشین‌بینایی و پردازش تصویر*. شماره ۴، ص ۱۷-۲۸، ۱۳۹۶

- IEEE International Workshop on Information Forensics and Security*, Rome, Italy, November 16–19 2015.
- [41] T. Denemark and J. Fridrich, "Model based steganography with precover," In A. Alattar and N. D. Memon, editors, *Proceedings IS&T, Electronic Imaging, Media Watermarking, Security, and Forensics 2017*, San Francisco, CA, January 29–February 1, 2017.
- [42] T. Denemark, "Side-Information For Steganography Design And Detection", State University of New York at Binghamton, 2018.
- [43] Boroumand, M., Fridrich, J., "Synchronizing embedding changes in side-informed steganography", *Electronic Imaging*, vol 2020, no 4, bll 290–291, 2020
- [44] W. Su, J. Ni, X. Li, and Y. Q. Shi, "A new distortion function design for jpeg steganography using the generalized uniform embedding strategy," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 12, pp. 3545–3549, 2018.
- [45] L. Guo, J. Ni, W. Su, C. Tang, and Y. Q. Shi, "Using statistical image model for JPEG steganography: Uniform embedding revisited," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 12, pp. 2669–2680, 2015.
- [46] X. Hu, J. Ni, and Y. Q. Shi, "Efficient jpeg steganography using domain transformation of embedding entropy," *IEEE Signal Processing Letters*, vol. 25, no. 6, pp. 773–777, 2018.
- [47] K. Chen, H. Zhou, W. Zhou, W. Zhang, and N. Yu, "Defining cost functions for adaptive jpeg steganography at the microscale," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 4, pp. 1052–1066, 2019.
- [48] X. Liao, Y. Yu, B. Li, Z. Li, en Z. Qin, "A new payload partition strategy in color image steganography", *IEEE Transactions on Circuits and Systems for Video Technology*, vol 30, no 3, bll 685–696, 2019.
- [49] Q. Giboulot, R. Cogranne, and P. Bas. Synchronization Minimizing Statistical Detectability for Side-Informed JPEG Steganography. In *IEEE International Workshop on Information Forensics and Security*, New York, NY, December 6–11, 2020.
- [50] J. Butora and J. Fridrich. Steganography and its detection in JPEG images obtained with the "trunc" quantizer. In *Proceedings IEEE, International Conference on Acoustics, Speech, and Signal Processing*, Barcelona, Spain, May 4–8, 2020.
- [51] R. Cogranne, Q. Giboulot, en P. Bas, "Efficient Steganography in JPEG Images by Minimizing Performance of Optimal Detector", *IEEE Transactions on Information Forensics and Security*, 2021.
- [52] Filler, T., Judas, J., and Fridrich, J., "Minimizing additive distortion in steganography using syndrometrellis codes," on information forensics and security (WIFS), pp.234–239, IEEE, 2012.
- [29] Holub, V., Fridrich, J., and Denemark, T., "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol.2014, no.1, p.1, 2014.
- [30] Fridrich, J. and Kodovský, J., "Multivariate gaussian model for designing additive distortion for steganography," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.2949–2953, IEEE, 2013.
- [31] Sedighi, V., Cogranne, R., and Fridrich, J., "Content adaptive steganography by minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security*, vol.11, no.2, pp.221–234, 2015.
- [32] Sedighi, V., Fridrich, J., and Cogranne, R., "Content adaptive pentary steganography using the multivariate generalized Gaussian cover model," in *Media Watermarking, Security, and Forensics 2015*, vol.9409, pp.144–156, International Society for Optics and Photonics, 2015.
- [33] Li, B., Wang, M., Huang, J., and Li, X., "A new cost function for spatial image steganography," in *2014 IEEE International Conference on Image Processing (ICIP)*, pp.4206–4210, IEEE, 2014.
- [34] Li, B., Wang, M., Li, X., Tan, S., and Huang, J., "A strategy of clustering modification directions in spatial image steganography," *IEEE Transactions on Information Forensics and Security*, vol.10, no.9, pp.1905–1917, 2015.
- [35] Denemark, T. and Fridrich, J., "Improving steganographic security by synchronizing the selection channel," in *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*, pp.5–14, ACM, 2015.
- [36] Zhang, W., Zhang, Z., Zhang, L., Li, H., and Yu, N., "Decomposing joint distortion for adaptive steganography," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.27, no.10, pp.2274–2280, 2016.
- [37] Su, W., Ni, J., Hu, X., and Fridrich, J., "Image steganography with symmetric embedding using gaussian markov random field model," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.
- [۳۸] عبدالهی، بهناز، هراتی، احد، طاهری‌نیا، امیرحسین. "پنهان‌نگاری متقارن مبتنی بر استنتاج میدان متوسط"، علوم رایانش و فناوری اطلاعات، ۱۴۰۱.
- [39] B. Abdollahi, A. Harati, and A. Taherinia, "Non-additive image steganographic framework based on variational inference in Markov Random Fields," *Journal of Information Security and Applications*, vol. 68, p. 103254, 2022.
- [40] Denemark, T., Sedighi, V., Holub, V., Cogranne, R., and Fridrich, J., "Side-informed steganography with additive distortion," In

- incomplete cover model,” in *Proceedings of the thirteenth ACM multimedia workshop on Multimedia and security*, pp.69–76, 2011.
- [66] Kouider, S., Chaumont, M., and Puech, W., “Adaptive steganography by oracle (aso),” in *2013 IEEE International Conference on Multimedia and Expo (ICME)*, pp.1–6, IEEE, 2013.
- [67] Zhang, K. A., CuestaInfante, A., Xu, L., and Veeramachaneni, K., “Steganogan: high capacity image steganography with gans,” *arXiv preprint arXiv:1901.03892*, 2019.
- [68] Yedroudj, M., Comby, F., and Chaumont, M., “Steganography using a 3player game,” *Journal of Visual Communication and Image Representation*, p.102910, 2020.
- [69] W. Shi and S. Liu, “Hiding Message Using a Cycle Generative Adversarial Network,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2022.
- [70] Bas, P., Filler, T., and Pevný, T., “break our steganographic system”: the ins and outs of organizing boss,” in *International workshop on information hiding*, pp.59–70, Springer, 2011.
- [71] Denemark, T., Sedighi, V., Holub, V., Coganne, R., and Fridrich, J., “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, 2014.
- [72] Ker, A. D., Pevny, T., and Bas, P., “Rethinking optimal embedding,” in *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*, pp.93–102, 2016.
- [73] Butora, J., Yousfi, Y., and Fridrich, J., “Turning cost based steganography into model based,” in *Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security*, pp.151–159, 2020.



بهناز عبدالمی دانش‌آموخته دکترای هوش مصنوعی در سال ۱۴۰۱ از دانشگاه فردوسی مشهد است. وی مدرک کارشناسی را در رشته مهندسی کامپیوتر از دانشگاه شهید باهنر کرمان و کارشناسی ارشد نرم‌افزار را از دانشگاه صنعتی اصفهان به ترتیب در سال‌های ۱۳۸۷ و ۱۳۹۰ دریافت کرد. علایق پژوهشی او شامل مدل‌سازی احتمالاتی، پردازش تصویر، نهان‌نگاری و یادگیری عمیق است.

نشانی رایانامه ایشان، عبارت است از:

b.abdollahi@mail.um.ac.ir

- IEEE Transactions on Information Forensics and Security*, vol.6, no.3, pp.920–935, 2011.
- [53] Pevny, T., Bas, P., and Fridrich, J., “Steganalysis by subtractive pixel adjacency matrix,” *IEEE Transactions on Information Forensics and Security*, vol.5, no.2, pp.215–224, 2010.
- [54] Fridrich, J. and Kodovsky, J., “Rich models for steganalysis of digital images,” *IEEE Transactions on Information Forensics and Security*, vol.7, no.3, pp.868–882, 2012.
- [55] Goodfellow, I., PougetAbadie, J., Mirza, M., Xu, B., WardeFarley, D., Ozair, S., Courville, A., and Bengio, Y., “Generative adversarial nets,” in *Advances in neural information processing systems*, pp.2672–2680, 2014.
- [56] Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R., “Intriguing properties of neural networks,” *arXiv preprint arXiv:1312.6199*, 2013.
- [57] Zhang, Y., Zhang, W., Chen, K., Liu, J., Liu, Y., and Yu, N., “Adversarial examples against deep neural network based steganalysis,” in *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security*, pp.67–72, 2018.
- [58] Tang, W., Li, B., Tan, S., Barni, M., and Huang, J., “Cnn based adversarial embedding for image steganography,” *IEEE Transactions on Information Forensics and Security*, vol.14, no.8, pp.2074–2087, 2019.
- [59] Ma, S., Zhao, X., and Liu, Y., “Adaptive spatial steganography based on adversarial examples,” *Multimedia Tools and Applications*, vol.78, no.22, pp.32503–32522, 2019.
- [60] Liu, M., Song, T., Luo, W., Zheng, P. and Huang, J., “Adversarial steganography embedding via stego generation and selection” *IEEE Transactions on Dependable and Secure Computing*, 2022.
- [61] Qin, X., Li, B., Tan, S., Tang, W., Huang, J., “Gradually Enhanced Adversarial Perturbations on Color Pixel Vectors for Image Steganography”, *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.
- [62] Tang, W., Tan, S., Li, B., and Huang, J., “Automatic steganographic distortion learning using a generative adversarial network,” *IEEE Signal Processing Letters*, vol.24, no.10, pp.1547–1551, 2017.
- [63] Yang, J., Ruan, D., Huang, J., Kang, X., and Shi, Y.Q., “An embedding cost learning framework using gan,” *IEEE Transactions on Information Forensics and Security*, vol.15, pp.839–851, 2019.
- [64] Li, F., Yu, Z. and Qin, C., “GAN-based spatial image steganography with cross feedback mechanism” *Signal Processing*, 190, p.108341, 2022.
- [65] Kodovsky, J., Fridrich, J., and Holub, V., “On dangers of overtraining steganography to



احد هراتی در سال ۱۳۸۲ مدرک کارشناسی ارشد هوش مصنوعی و رباتیک را از دانشگاه تهران و در سال ۱۳۸۷ دکترای خود را از موسسه فناوری فدرال زوریخ (ETHZ)، زوریخ، سوئیس دریافت کرد. وی در حال حاضر به عنوان دانشیار گروه مهندسی کامپیوتر، دانشگاه فردوسی مشهد در حال فعالیت است. علایق تحقیقاتی او شامل ادراک ربات به ویژه دید سه بعدی، یادگیری تقویتی و مدل‌های احتمالاتی است.

نشانی رایانامه ایشان، عبارت است از:
a.harati@um.ac.ir



امیرحسین طاهری‌نیا مدرک کارشناسی خود را سال ۱۳۸۳ در رشته مهندسی کامپیوتر از دانشگاه فردوسی مشهد و کارشناسی ارشد و دکترای مهندسی کامپیوتر را در سال‌های ۱۳۸۵ و ۱۳۹۰ از دانشگاه صنعتی شریف دریافت کرد. وی در حال حاضر به عنوان دانشیار گروه مهندسی کامپیوتر، دانشگاه فردوسی مشهد در حال فعالیت است. علایق پژوهشی او شامل امنیت چندرسانه‌ای، پنهان‌سازی داده‌ها و پردازش سیگنال چندرسانه‌ای است.

نشانی رایانامه ایشان، عبارت است از: