Estudo sobre Ferramentas de Software Livre para Detecção de Esteganografia em Imagens Digitais

Érico Meger[1], **Eros Henrique Lunardon Andrade**[1], **Guilherme Werneck de Oliveira**[1]

[1]Campus Pinhais – Instituto Federal do Paraná (IFPR) Pinhais - PR - Brasil

Abstract. *This work analyzes free and open-source tools for image steganography and steganalysis, covering both traditional methods and modern approaches based on convolutional neural networks (CNNs). Classical tools such as Aletheia and StegExpose provide accessibility and transparency, yet show limitations when facing advanced steganographic algorithms. Conversely, CNN-based models including SRNet, EfficientNet, and MixNet achieve superior performance by automatically learning subtle modification patterns but require significant computational resources. The analysis reveals a paradigm shift in the field and highlights the importance of open-source software in enabling research. The study concludes that future steganalysis solutions must balance accuracy, efficiency, and usability.*

Resumo. *Este trabalho analisa ferramentas de software livre voltadas à esteganografia e estegonálise em imagens digitais, abrangendo métodos tradicionais e abordagens modernas baseadas em redes neurais convolucionais (CNNs). Ferramentas clássicas, como Aletheia e StegExpose, oferecem acessibilidade e transparência, porém apresentam limitações frente a algoritmos esteganográficos avançados. Já as CNNs como SRNet, EfficientNet e MixNet demonstram desempenho superior ao aprenderem automaticamente padrões sutis de modificação, embora dependam de alto poder computacional. A análise evidencia uma transição de paradigma na área e destaca o papel do software livre na democratização da pesquisa. Conclui-se que o futuro da estegonálise requer soluções que combinem precisão, eficiência e usabilidade.*

1. Introdução

O movimento do software livre se estabelece como um paradigma essencial para promover transparência, colaboração e inovação no cenário tecnológico contemporâneo. Segundo a Free Software Foundation, software livre é definido pela sua capacidade de respeitar as liberdades e o controle dos usuários sobre o software: a liberdade de executar o programa para qualquer propósito, de estudá-lo e modificá-lo (acesso ao código-fonte é pré-requisito), de redistribuir cópias e de distribuir versões modificadas para a comunidade, conhecidas como as quatro liberdades essenciais [Foundation 2024].

Ao assegurar essas liberdades, o software livre não apenas fortalece a confiança nas soluções digitais, por permitir auditoria e aprendizado mútuo, mas também fomenta ambientes colaborativos dinâmicos, onde ferramentas podem ser aprimoradas coletivamente.

No contexto da esteganografia, a disponibilidade dessas ferramentas open source oferece grandes oportunidades para o avanço da área.

A análise de imagens digitais, por exemplo, pode se beneficiar de recursos de detecção de padrões e classificação automática fornecidos por essas ferramentas, auxiliando tanto no desenvolvimento de técnicas esteganográficas mais robustas quanto na criação de métodos de detecção mais eficazes. Assim, a intersecção entre software livre, inteligência artificial e esteganografia evidencia como a filosofia do código aberto não só fortalece a confiança técnica, mas também amplia as possibilidades de pesquisa e aplicação prática neste campo.

A esteganografia pode ser compreendida como uma técnica utilizada para esconder informações em meios aparentemente comuns, de forma que um observador externo não consiga identificar a presença de dados ocultos [Fridrich 2010].

Essa área de estudo, portanto, não se limita apenas ao ato de esconder informações, mas constitui um campo de estudo mais amplo que abrange técnicas, algoritmos e aplicações destinadas a garantir a confidencialidade e a discrição da comunicação. Em contraste com a criptografia, que protege o conteúdo das mensagens mas não oculta sua existência, a esteganografia busca mascarar o próprio ato de comunicação [Fridrich 2010]. Essa característica a torna uma área estratégica tanto para aplicações legítimas, como autenticação de documentos e proteção da privacidade, quanto para usos maliciosos. Tal dualidade evidencia que a esteganografia deve ser compreendida não apenas sob uma perspectiva técnica, mas também dentro de um contexto social e político mais amplo.

Nesse sentido, ao longo da história, e de forma ainda mais acentuada no cenário contemporâneo, observa-se o fortalecimento de mecanismos de vigilância e controle sobre a comunicação digital. Na Europa, por exemplo, esse movimento se materializa tanto em iniciativas de remoção massiva de conteúdos, com mais de 41 milhões de postagens bloqueadas apenas no primeiro semestre de 2025 [Poder360 2025], quanto em pressões políticas para enfraquecer a segurança criptográfica, como a exigência de um backdoor no iCloud, que levou a Apple a retirar a opção de criptografia de ponta a ponta de seus serviços no Reino Unido [Guardian 2025]. Embora tais medidas sejam frequentemente justificadas em nome da segurança pública, a ausência de transparência sobre os critérios de censura e o impacto direto na privacidade digital levantam sérias preocupações. Nesse contexto, a esteganografia age como uma alternativa tecnológica de resistência, capaz de proporcionar meios de comunicação discretos e seguros, reforçando sua relevância sociopolítica e justificando o aprofundamento de seu estudo.

1.1. Objetivo

Analisar ferramentas de software livre voltadas à esteganografia e à esteganálise em imagens digitais, considerando tanto soluções tradicionais quanto abordagens recentes baseadas em CNNs. Busca-se compreender como essas ferramentas se estruturam, quais funcionalidades oferecem e de que forma contribuem para o avanço da área, identificando tendências, limitações e possibilidades de integração entre métodos clássicos e modernos.

2. Revisão bibliográfica

Esta seção apresenta uma revisão das principais ferramentas de software livre desenvolvidas para a detecção de esteganografia em imagens digitais, descrevendo suas abordagens, funcionalidades e limitações. Além dessas ferramentas tradicionais, também

são consideradas soluções baseadas em redes neurais convolucionais (CNNs) de código aberto, como a EfficientNet, que atualmente representam o estado da arte na estegoanálise [La Croix et al. 2024].

O trabalho *"An Ensemble Model using CNNs on Different Domains for ALASKA2 Image Steganalysis"* de Chubachi [Chubachi 2020] surge da constatação de que muitos detectores de esteganografia baseados em aprendizado profundo não generalizam bem em cenários reais devido ao uso de conjuntos de dados simplificados. A competição ALASKA2 ofereceu um ambiente mais realista, com imagens JPEG coloridas de diferentes origens e processos, estimulando soluções mais aplicáveis. Nesse contexto, o objetivo do autor foi desenvolver um modelo de detecção baseado em um ensemble de redes convolucionais que combinasse informações tanto do domínio espacial (RGB, YUV e Lab) quanto do domínio da frequência (coeficientes DCT).

A metodologia proposta envolveu CNNs construídas sobre arquiteturas EfficientNet, com ajustes para lidar com as especificidades de cada domínio. No caso dos coeficientes DCT, foram aplicadas codificações one-hot, recortes de valores e convoluções dilatadas para capturar padrões. Para integrar os modelos, além da simples média de previsões, foi desenvolvido um perceptron multicamada capaz de combinar os mapas de características. Também se utilizaram técnicas auxiliares, como pseudo-rotulagem e stacking com LightGBM. Em experimentos conduzidos com 300 mil imagens, o uso combinado dos modelos trouxe ganhos consistentes, resultando em uma performance de AUC ponderado próxima de 0,94 e garantindo a terceira colocação na competição.

O estudo apresenta como pontos fortes a inovação de combinar diferentes domínios e a validação em um cenário competitivo e realista. Contudo, o alto custo computacional e a limitação de testar apenas algoritmos de esteganografia já conhecidos restringem sua aplicabilidade prática.

Explorando ensembles de forma mais sistemática, o estudo *Ensemble of CNNs for Steganalysis: An Empirical Study* de Xu et al. [Xu et al. 2016] retorna ao tema de ensembles, mas com uma abordagem mais empírica. Motivado pela observação de que, embora as Redes Neurais Convolucionais (CNNs) estivessem ganhando popularidade em esteganálise, a maioria das pesquisas se concentrava no design de um único modelo de CNN. No entanto, no campo mais amplo da visão computacional e do aprendizado de máquina, as melhores performances são consistentemente alcançadas por meio de ensembles, ou seja, a combinação de múltiplos modelos. Os autores perceberam uma lacuna na literatura de esteganálise, que ainda não havia explorado a fundo estratégias de ensemble mais sofisticadas do que a simples média das previsões. O principal objetivo do trabalho foi, portanto, conduzir um estudo empírico para avaliar o desempenho de diferentes estratégias de combinação de CNNs para a tarefa de esteganálise, buscando ir além da média de modelos e testar o uso de classificadores de segundo nível treinados sobre as saídas e representações internas das redes.

A metodologia proposta envolveu, primeiramente, o treinamento de um conjunto de 16 CNNs individuais, que serviram como "aprendizes de base", cada uma treinada sobre um subconjunto aleatório do dataset de treinamento. A partir disso, os autores testaram e compararam três abordagens de ensemble: a média simples das probabilidades de saída, a criação de novos vetores de características a partir dessas probabilidades para

treinar um classificador de segundo nível, e uma terceira técnica mais inovadora que consistia em extrair as representações de características das camadas intermediárias de cada CNN, concatená-las e usá-las para treinar o classificador de segundo nível. O experimento foi realizado no dataset BOSSbase v1.01 para detectar a esteganografia do algoritmo S-UNIWARD com uma taxa de inserção de 0.4 bpp.

O trabalho apresenta como pontos fortes a demonstração sistemática de que o uso de um classificador de segundo nível sobre as saídas das CNNs melhora consistentemente o desempenho em relação à média de modelos, e, principalmente, a descoberta de que o uso das representações de características intermediárias resulta na melhor performance, indicando que os classificadores mais robustos podem extrair padrões mais discriminativos do que as camadas de classificação simples das CNNs base. As limitações do estudo, no entanto, incluem a sua validação em apenas um dataset, contra um único algoritmo esteganográfico e com uma única taxa de inserção, além do alto custo computacional da abordagem.

Complementando a abordagem de ensemble, o trabalho *ImageNet Pre-trained CNNs for JPEG Steganalysis* de Yousfi et al. [Fridrich et al. 2020] explora uma direção diferente: o uso de transfer learning em esteganálise. A principal motivação para o estudo surgiu a partir da competição de esteganálise ALASKA II, onde foi observado que os participantes com melhor desempenho utilizavam modelos de visão computacional de uso geral, como EfficientNet e ResNet, em vez de arquiteturas especializadas e treinadas do zero para a tarefa. Essa nova abordagem, baseada em aprendizagem por transferência (transfer learning), representava uma mudança de paradigma em relação a modelos consolidados, como a SRNet, que eram projetados especificamente para esteganálise. Diante desse cenário, o principal objetivo dos autores foi investigar e demonstrar formalmente a eficácia e a superioridade desses modelos pré-treinados no ImageNet para a detecção de esteganografia em imagens JPEG, comparando seu desempenho com as abordagens tradicionais.

A metodologia utilizada centrou-se em refinar (fine-tuning) diversas arquiteturas pré-treinadas, como EfficientNet, MixNet e ResNet, no conjunto de dados da ALASKA II. Os autores também conduziram experimentos para avaliar o impacto de decisões arquitetônicas, como a remoção de camadas de pooling ou stride no início da rede, confirmando que a manutenção da resolução original nas primeiras camadas é crucial para a performance em esteganálise. O experimento principal foi conduzido no dataset da ALASKA II, que continha imagens comprimidas com fatores de qualidade 75, 90 e 95 e com mensagens ocultas pelos algoritmos J-UNIWARD, J-MiPOD e UERD.

Os pontos fortes dessa abordagem, destacados no artigo, são a acurácia superior, a maior eficiência de dados e uma velocidade de treinamento ordens de magnitude mais rápida em comparação com o treinamento de um modelo do zero. Como limitação, os próprios autores apontam que o estudo foi amplamente focado no ambiente da ALASKA II e que o ganho de desempenho obtido com o pré-treinamento tende a diminuir à medida que o volume de dados para treinamento na tarefa final aumenta.

Diferentemente das abordagens anteriores que se concentram em arquiteturas de ensemble ou transfer learning, o artigo *An Intriguing Struggle of CNNs in JPEG Steganalysis and the OneHot Solution* de Yousfi e Fridrich [Yousfi e Fridrich 2020] identifica e re-

solve uma limitação específica das CNNs em esteganálise. A pesquisa parte da descoberta de cenários específicos onde as modernas CNNs, como a SRNet, apresentavam um desempenho surpreendentemente inferior ao de métodos mais antigos baseados em extração de características, como o JPEG Rich Model (JRM). Essa falha era particularmente evidente na detecção do algoritmo nsF5 e do J-UNIWARD em certos tipos de imagem JPEG, e a análise revelou que o sucesso do JRM nesses casos se devia à sua capacidade de computar estatísticas simples dos coeficientes DCT, como histogramas de coocorrência, algo que as CNNs convencionais, que operam na imagem descomprimida, não conseguiam ”enxergar”.

A metodologia se baseia em duas inovações principais. A primeira é a própria rede ”OneHot CNN”, que transforma sua entrada através de uma camada de ”codificação one-hot com corte” (clipped one-hot encoding). Nessa etapa, os valores absolutos dos coeficientes DCT são transformados em um volume binário que facilita o aprendizado de ocorrências e coocorrências pelas camadas convolucionais subsequentes. A segunda inovação é a arquitetura de ramo duplo ”OneHot+SRNet”, que mescla a nova rede OneHot com uma CNN convencional (SRNet). O experimento principal consistiu em testar o desempenho dessas novas arquiteturas nos cenários problemáticos (nsF5 e J-UNIWARD) em datasets como BOSSbase e BOWS2.

O trabalho se destaca por identificar com precisão uma falha em modelos estado da arte e propor uma solução elegante e eficaz, a codificação one-hot, que permite a uma CNN aprender estatísticas de alta ordem de forma flexível. A arquitetura de ramo duplo é outro ponto forte, pois oferece uma maneira prática de criar um detector mais completo e robusto. Uma limitação implícita é que a rede OneHot é altamente especializada para esses casos de falha, o que justifica sua fusão com uma rede mais geral como a SRNet.

Para contextualizar os trabalhos anteriores dentro do panorama geral da área, o artigo *Comprehensive survey on image steganalysis using deep learning* de De La Croix et al. [La Croix et al. 2024] oferece uma visão abrangente do estado da arte. A principal motivação dos autores reside na observação de que as abordagens tradicionais, baseadas em aprendizado de máquina (ML), se mostraram ineficazes contra os modernos algoritmos de esteganografia. Métodos clássicos como Support Vector Machines (SVM) e Ensemble Classifiers (EC) dependem de um processo árduo e manual de extração de características, o que não só consome tempo, mas também sofre com a ”maldição da dimensionalidade”, onde o excesso de características prejudica o desempenho do classificador. A introdução do aprendizado profundo (deep learning) marcou uma mudança de paradigma, unificando a extração de características e a classificação em um único processo otimizado e de ponta a ponta.

A metodologia empregada pelos autores é a de uma revisão sistemática da literatura, baseada no protocolo PRISMA. Eles selecionaram 24 artigos de ponta, publicados entre 2014 e 2023, que representam a vanguarda da esteganálise com deep learning. O artigo estrutura-se de forma didática, iniciando com uma taxonomia detalhada das técnicas de esteganálise, explorando a transição do paradigma de ML para o de deep learning, e analisando as arquiteturas de Redes Neurais Convolucionais (CNNs) propostas, desde as pioneiras até as mais recentes. A pesquisa documenta a progressão das arquiteturas, começando com modelos como Qian-Net, passando pelo Xu-Net, chegando a modelos mais sofisticados como o Ye-Net, e culminando em arquiteturas estado da arte como a

SRNet e a GBRAS-Net.

Um dos pontos fortes mais significativos desta revisão é a identificação dos principais desafios que ainda persistem no campo da esteganálise com deep learning: vulnerabilidade a ataques adversariais, qualidade e padronização dos datasets, ineficiência para lidar com imagens de tamanhos arbitrários, dificuldade na detecção de baixo payload, problema do cover-source mismatch, dificuldade na identificação de características globais, e necessidade de um grande número de amostras de treinamento. Como limitação, a revisão foca, que principalmente em imagens no domínio espacial, deixando lacunas para outros domínios como JPEG.

A pesquisa de Farooq e Selwal [Farooq e Selwal 2023] é motivada pela crescente complexidade e sofisticação tanto das técnicas de esteganografia quanto das de esteganálise. Os autores partem da premissa de que, com os avanços em comunicação e tecnologia da informação, os métodos para ocultar dados em imagens tornaram-se mais robustos, desafiando as abordagens de detecção tradicionais [Farooq e Selwal 2023]. Enquanto métodos clássicos se baseiam em um processo de duas etapas—extração manual de características (como os Rich Models) e classificação (usando SVMs ou Ensemble Classifiers)—, o aprendizado profundo (deep learning, DL) emergiu como uma alternativa superior, unificando essas etapas e aprendendo automaticamente as características mais relevantes diretamente dos dados brutos [Farooq e Selwal 2023]. Essa mudança de paradigma gerou uma vasta quantidade de novas arquiteturas e abordagens, criando a necessidade de uma análise consolidada. Assim, o principal objetivo do trabalho é realizar uma revisão sistemática e aprofundada das técnicas de esteganálise de imagens baseadas em DL, oferecendo um panorama do estado da arte, analisando comparativamente os modelos, os datasets de referência e as métricas de avaliação, além de identificar os desafios de pesquisa que permanecem abertos para guiar futuros investigadores no campo [Farooq e Selwal 2023].

A metodologia adotada no artigo é a de uma revisão sistemática e abrangente da literatura. Primeiramente, os autores fornecem uma classificação detalhada das técnicas de esteganálise, cobrindo desde ataques visuais e estatísticos até abordagens mais complexas como a análise estrutural e métodos baseados em redes neurais artificiais (ANN), estabelecendo um contexto histórico e técnico [Farooq e Selwal 2023]. O foco principal, no entanto, é a análise cronológica das abordagens baseadas em DL, com especial atenção às Redes Neurais Convolucionais (CNNs). O artigo revisa uma sequência de trabalhos influentes, começando com as primeiras propostas de Tan et al. (2014) e Qian et al. (2015), e avançando para arquiteturas mais complexas e eficazes como as de Xu et al. (2016), Ye et al. (2017), Yedroudj et al. (2018), e a SRNet de Boroumand et al. (2019) [Farooq e Selwal 2023]. Para cada trabalho, os autores analisam a arquitetura da rede, as funções de ativação utilizadas, as bases de dados (como BOSSBase, BOWS2 e ImageNet), os algoritmos esteganográficos alvo (S-UNIWARD, WOW, HILL, etc.) e os resultados de desempenho reportados. Além disso, a pesquisa compila e discute os principais datasets de referência, as métricas de avaliação quantitativas (como taxa de erro, curva ROC, AUC e WAUC) e qualitativas (PSNR, SSIM), e até mesmo lista ferramentas de software de código aberto e proprietárias disponíveis para esteganálise [Farooq e Selwal 2023].

Como um artigo de revisão, o experimento consiste na compilação e na análise

crítica dos resultados publicados por outros pesquisadores. A pesquisa consolida os dados de desempenho de diversas arquiteturas de CNN em uma tabela comparativa detalhada, avaliando a taxa de erro de detecção contra diferentes algoritmos esteganográficos (como HUGO, WOW, S-UNIWARD) e em diferentes taxas de inserção de dados (payloads) [Farooq e Selwal 2023]. A análise dos resultados mostra uma clara evolução: os primeiros modelos de CNN apresentavam desempenho comparável ou ligeiramente inferior aos métodos tradicionais (como o SRM com Ensemble Classifier), mas com o avanço no design das arquiteturas—incorporando camadas de pré-processamento com filtros, funções de ativação mais adequadas (como a TLU), normalização em lote (BN), e arquiteturas residuais profundas—, os modelos mais recentes superaram significativamente os métodos clássicos, especialmente em cenários com maior payload [Farooq e Selwal 2023]. O artigo destaca, por exemplo, o desempenho superior de modelos como o de Yedroudj et al. e a SRNet, que estabeleceram novos padrões de acurácia na detecção [Farooq e Selwal 2023].

Os pontos fortes deste artigo são a sua abrangência e o seu caráter sistemático. Ele funciona como um documento de referência completo, não apenas listando os trabalhos, mas contextualizando-os, explicando a evolução das ideias e das arquiteturas, e oferecendo uma análise comparativa clara. Outro ponto forte crucial é a identificação detalhada das limitações e dos desafios de pesquisa em aberto, que incluem: a dificuldade na detecção de características estatísticas de sinais esteganográficos fracos; o desafio persistente da detecção em cenários de baixo payload; os problemas de generalização (como steganographic mismatch, cover-source mismatch e payload mismatch); e a necessidade de mais pesquisas em esteganálise quantitativa e locativa usando DL [Farooq e Selwal 2023]. Uma limitação inerente a qualquer trabalho de revisão é que ele depende da qualidade e da consistência dos dados reportados pelos artigos originais, não gerando novos resultados empíricos por si só.

Buscando avançar o estado da arte em esteganálise no domínio espacial, o artigo *A convolutional neural network to detect possible hidden data in spatial domain images* de La Croix [La Croix e Ahmad 2023] propõe uma nova arquitetura de rede neural convolucional. A principal motivação dos autores é a observação de que, embora as CNNs existentes tenham superado as técnicas clássicas, elas ainda apresentam problemas de desempenho em termos de acurácia de classificação e estabilidade durante o treinamento. O objetivo do trabalho é, portanto, projetar uma nova arquitetura que enderece essas lacunas, melhorando tanto a precisão na detecção de dados ocultos quanto a estabilidade do treinamento, através de inovações específicas nas fases de pré-processamento, extração de características e classificação.

A metodologia consiste no desenvolvimento de uma CNN customizada que se inicia com filtros não treináveis do Spatial Rich Model (SRM) para amplificar o ruído esteganográfico. A principal inovação está na fase de extração de características, onde se utiliza uma combinação de convoluções separáveis em profundidade (depthwise separable convolutions) com a função de ativação Leaky Rectified Linear Unit (LReLU) para reduzir parâmetros, evitar o sobreajuste e melhorar a estabilidade do gradiente. Na classificação, um módulo de pooling multi-escala agrega as características de forma eficiente. O experimento principal comparou o desempenho da rede proposta com modelos de ponta (Ye-Net, GBRAS-Net) nos datasets BOSSbase e BOWS 2 contra os algoritmos

S-UNIWARD e WOW, demonstrando consistentemente uma performance superior.

Um dos pontos fortes mais significativos do trabalho é o ganho de desempenho expressivo sobre o estado da arte, com melhorias na acurácia que chegam a mais de 10% em alguns cenários. Os autores validam rigorosamente a contribuição de cada componente inovador através de um estudo de ablação, que demonstra que o uso de convoluções separáveis, a ativação LReLU e o pooling multi-escala, individualmente, trazem ganhos substanciais para o resultado final. Como limitação, a avaliação do artigo se concentra exclusivamente em algoritmos do domínio espacial, não explorando o desempenho da arquitetura em domínios de transformação como o JPEG, e embora apresente resultados preliminares para imagens de tamanhos arbitrários, uma análise aprofundada desse cenário é declarada como fora do escopo do trabalho.

Em contraste com os trabalhos revisados, que focam em ....., a proposta desenvolvida neste trabalho busca explorar o uso de bibliotecas de software livre para a construção de modelos de inteligência artificial em esteganálise, democratizando a área ao garantir reprodutibilidade, baixo custo e acessibilidade.

3. Metodologia

A pesquisa será conduzida de forma qualitativa e descritiva. Inicialmente, será realizada uma revisão bibliográfica sobre os fundamentos da esteganografia e da esteganálise, com ênfase em métodos e algoritmos empregados na detecção de dados ocultos em imagens digitais. Em seguida, serão mapeadas e analisadas ferramentas de software livre representativas dessas duas abordagens: ferramentas que oferecem recursos práticos de inserção e detecção de mensagens, e soluções recentes baseadas em aprendizado profundo, que utilizam redes neurais convolucionais para aprimorar a acurácia da detecção. A análise considerará aspectos como objetivos das ferramentas, funcionalidades disponibilizadas, técnicas implementadas, documentação, acessibilidade e relevância no contexto atual da pesquisa em esteganálise. Por fim, será realizada uma discussão comparativa entre os diferentes tipos de ferramentas, destacando suas complementaridades e os caminhos evolutivos da área.

4. Resultados e Discussões

A análise foi conduzida a partir de um levantamento de ferramentas de software livre voltadas à esteganografia e à esteganálise em imagens digitais. As ferramentas foram agrupadas em duas categorias principais: (i) soluções tradicionais, baseadas em métodos clássicos de inserção e detecção de mensagens ocultas, e (ii) soluções modernas, baseadas em redes neurais convolucionais (CNNs), que representam o atual estado da arte na detecção automática de esteganografia. Os critérios de análise consideraram aspectos como objetivo, abordagem técnica, tipo de licença, disponibilidade de código, documentação, acessibilidade e relevância para pesquisa e desenvolvimento.

4.1. Ferramentas tradicionais

As ferramentas tradicionais representam a base histórica da esteganografia prática, fornecendo utilitários acessíveis para detecção de dados ocultos.

Tabela 1. Ferramentas tradicionais de software livre para estegoanálise.

Ferramenta	Objetivo	Abordagem Técnica	Licença
Aletheia [Hostalot e Megías 2024]	Estegoanálise estatística e visual	Análise de ruído, histogramas e artefatos de compressão	MIT
StegExpose [Boehm 2014]	Detecção automatizada de esteganografia em lote	Combina métricas como Sample Pair e RS Analysis	Não especificada

A Tabela 1 apresenta um resumo comparativo das principais soluções software livre de estegoanálise em imagem analisadas.

A Aletheia revela seu duplo propósito como uma ferramenta para analistas forenses e pesquisadores. Suas funcionalidades vão além da simples detecção, incorporando ataques estruturais a formatos de imagem, análise de metadados e, principalmente, a capacidade de realizar análises visuais e estatísticas. Por exemplo, a ferramenta permite a extração de histogramas de cores e a aplicação de filtros high-pass para amplificar ruídos, auxiliando um analista a identificar visualmente anomalias que podem indicar a presença de dados ocultos. Um diferencial importante da Aletheia é sua capacidade de lidar com o desafio do Cover Source Mismatch (CSM), um cenário comum em investigações do mundo real onde o modelo de detecção não foi treinado com o mesmo tipo de imagem que está sendo analisada. No entanto, sua principal limitação para usuários não especializados é a dependência da interpretação manual dos resultados, exigindo conhecimento técnico para tirar conclusões eficazes [Hostalot e Megías 2024].

O StegExpose, por sua vez, foi projetado com outro objetivo: a detecção automatizada e em lote de esteganografia LSB em imagens de formato lossless (sem perdas), como PNG e BMP. Sua principal inovação técnica é o uso de fusion techniques (técnicas de fusão), que combinam os resultados de múltiplos detectores estatísticos conhecidos, como RS Analysis[1] e Sample Pair Analysis[2], para produzir uma classificação final mais precisa do que cada método individualmente. A ferramenta foi desenvolvida com foco em praticidade e velocidade, oferecendo dois modos de operação: um modo ”padrão”, que maximiza a precisão, e um modo ”rápido”, que otimiza o tempo de análise ao descartar arquivos considerados ”limpos” nos estágios iniciais do processo. Sua principal limitação, reconhecida no próprio trabalho de origem, é ser especializada apenas em esteganografia LSB, não sendo projetada para detectar métodos mais avançados que operam no domínio da frequência (como em arquivos JPEG) [Boehm 2014].

Em comum, ambas ferramentas têm como principal mérito a acessibilidade e a transparência, características garantidas pela Aletheia, por conta de sua licença de software livre e pelo StegExpose que, apesar de não definir uma licença específica, é um projeto de código aberto. Além disso, ambas ferramentas contribuem para o ensino e

[1]A análise RS (Regular/Singular groups) explora como a inserção de dados LSB afeta a contagem de grupos de pixels regulares e singulares, revelando anomalias estatísticas.

[2]A Sample Pair Analysis examina pares de pixels e como seus valores mudam com a inserção de dados, detectando desvios em relação a uma imagem limpa.

a experimentação prática, permitindo a inspeção de algoritmos e a reprodutibilidade de resultados. No entanto, sua precisão na detecção é limitada, principalmente quando confrontadas com métodos modernos baseados em aprendizado profundo.

4.2. Abordagens baseadas em redes neurais convolucionais (CNNs)

Com o avanço da inteligência artificial, modelos de aprendizado profundo passaram a ser adotados para estegoanálise, com desempenho superior às abordagens anteriores [Boroumand et al. 2019].

Tabela 2. Redes neurais de código aberto aplicadas à estegoanálise.

Modelo	Arquitetura Base	Aplicação em Estegoanálise	Licença
SRNet [Boroumand et al. 2019]	CNN profunda especializada em ruído residual	Detecção de padrões sutis de modificação em pixels	MIT
EfficientNet [Tan e Le 2019a, Yousfi et al. 2021]	CNN escalável baseada em compound scaling	Fine-tuning em datasets como ALASKA II, com alto desempenho	Apache-2.0
MixNet [Tan e Le 2019b]	Variante da EfficientNet com convoluções mistas	Melhor generalização e leveza em comparação com EfficientNet	Apache-2.0

A Tabela 2 apresenta uma síntese das principais redes utilizadas na área, destacando suas licenças e relevância como ferramentas de pesquisa. Esses modelos representam uma evolução metodológica na área: ao contrário das ferramentas tradicionais, que dependem de heurísticas e análise manual, as CNNs aprendem a identificar padrões de esteganografia diretamente dos dados.

A SRNet (Steganalysis Residual Network), por exemplo, foi um marco por ser uma das primeiras arquiteturas profundas projetadas especificamente para estegoanálise. Sua principal inovação foi a criação de um bloco de camadas convolucionais sem pooling (agrupamento) e com conexões residuais, projetado para extrair e amplificar o ”ruído residual”, traços deixados pela inserção de dados, sem suprimir esse sinal fraco, algo que arquiteturas de classificação de imagem convencionais não fazem eficientemente. Ao ser ”agnóstica” a heurísticas, a SRNet demonstrou a superioridade do aprendizado de características de ponta a ponta (end-to-end) e se tornou uma baseline essencial para a comparação de novos modelos na área [Boroumand et al. 2019].

Posteriormente, a EfficientNet introduziu uma nova abordagem, não para a estegoanálise, mas para a eficiência de CNNs em geral. Sua arquitetura, encontrada por meio de busca de arquitetura neural (NAS), propõe um método de escalonamento composto (compound scaling) que equilibra de forma otimizada a profundidade, a largura e a resolução da rede. Pesquisadores em estegoanálise rapidamente adaptaram essa ideia, aplicando fine-tuning em modelos EfficientNet pré-treinados em grandes datasets como o ImageNet. Essa técnica provou ser altamente eficaz, alcançando ótimo desempenho em

benchmarks de estegoanálise como o ALASKA II, demonstrando que uma arquitetura eficiente para visão computacional geral também pode ser uma poderosa ferramenta para detectar padrões esteganográficos [Tan e Le 2019a].

Derivada da EfficientNet, a MixNet aprimora ainda mais a eficiência ao introduzir a Mixed Depthwise Convolution (MixConv). Em vez de usar um único tamanho de kernel convolucional por camada (ex: 3x3 ou 5x5), a MixConv combina múltiplos tamanhos de kernel em uma única operação. Essa abordagem permite que o modelo capture padrões em diferentes escalas e resoluções simultaneamente, resultando em melhor precisão e generalização com um custo computacional ainda menor. Para a estegoanálise, isso representa uma direção promissora para o desenvolvimento de modelos que sejam não apenas precisos, mas também leves o suficiente para serem integrados em aplicações práticas e com recursos limitados [Tan e Le 2019b].

No entanto, durante o levantamento bibliográfico realizado para esta pesquisa, notou-se que a MixNet foi a arquitetura menos frequentemente explorada.

4.3. Síntese comparativa e tendências

A análise das ferramentas mostra uma transição de paradigma na estegoanálise: uma passagem de métodos estatísticos, dependentes de interpretação humana, para abordagens automatizadas baseadas em aprendizado profundo. De um lado, ferramentas tradicionais como a Aletheia e o StegExpose democratizam o acesso à estegoanálise com foco em usabilidade e análise forense direcionada. A Aletheia, com suas análises visuais, serve como uma excelente ferramenta didática e investigativa, enquanto o StegExpose, com sua detecção em lote, oferece uma solução prática para varreduras em larga escala contra ataques LSB. Ambas, no entanto, atingem um teto de precisão por dependerem de heurísticas e características pré-definidas.

Do outro lado, modelos de CNN como a SRNet e adaptações da EfficientNet representam o estado da arte em detecção. A SRNet inaugurou a era do aprendizado de características de ponta a ponta, provando que uma rede pode aprender a ”ver” o ruído esteganográfico de forma mais eficaz que qualquer engenharia de características manual. A subsequente adoção da EfficientNet demonstra uma segunda tendência: a busca por eficiência. Contudo, essa evolução metodológica impõe uma nova barreira de entrada. O treinamento de tais modelos exige não apenas vastos conjuntos de dados, mas também um poder computacional significativo, envolvendo hardware especializado (como GPUs ou TPUs) e longos períodos de processamento. Esse requisito de recursos torna a replicação e o desenvolvimento de novas arquiteturas um desafio considerável para pesquisadores e desenvolvedores com acesso limitado a infraestrutura de alto desempenho, restringindo, em certo grau, a mesma democratização que as ferramentas clássicas promoveram.

O princípio do software livre é um fator comum e essencial para o avanço de ambas as abordagens. Nas abordagens tradicionais, ele garante transparência e acessibilidade. Nas abordagens modernas, seu papel é ainda mais crucial. É, que o caráter aberto de frameworks como TensorFlow e PyTorch que viabiliza a implementação, o compartilhamento e a rápida iteração sobre arquiteturas como as analisadas. Essa cultura de colaboração permite que a comunidade científica construa sobre o trabalho alheio, adaptando uma EfficientNet para estegoanálise ou propondo uma MixNet, acelerando o ciclo de inovação de uma forma que seria difícil de ver em um ecossistema de código fechado.

Por fim, a análise aponta para desafios persistentes e tendências futuras. A principal lacuna identificada é a ausência de ferramentas que integrem a alta precisão das CNNs com a usabilidade das soluções clássicas. Enquanto modelos como a MixNet apontam para um futuro de detectores leves e eficientes, eles permanecem, em grande parte, no domínio acadêmico. O próximo passo evolutivo para a estegoanálise de software livre parece ser, portanto, o desenvolvimento de soluções que unam esses dois mundos, criando ferramentas que sejam, ao mesmo tempo, poderosas, acessíveis e práticas para um público mais amplo.

5. Conclusão

O presente estudo atingiu seu objetivo de analisar ferramentas de software livre voltadas para a esteganografia e esteganálise, compreendendo tanto as soluções tradicionais quanto as baseadas no estado da arte em Redes Neurais Convolucionais.

A análise comparativa evidenciou uma transição de paradigma na área. Por um lado, as ferramentas tradicionais, como Aletheia e StegExpose, cumprem um papel fundamental na democratização do conhecimento, oferecendo interfaces acessíveis e transparência algorítmica essenciais para o ensino e para a análise forense preliminar. Contudo, sua dependência de heurísticas manuais impõe um teto de performance quando comparado a abordagens baseadas em redes neurais.

Por outro lado, as abordagens baseadas em CNNs, exemplificadas pela SRNet e pelas adaptações da EfficientNet, demonstraram superioridade técnica, aprendendo de forma automatizada. Entretanto, observou-se que essa evolução técnica trouxe consigo um novo desafio: a barreira do custo computacional. A exigência de hardware de alto desempenho para o treinamento desses modelos restringe a acessibilidade que o software livre visa promover, criando um distanciamento entre a precisão acadêmica e a aplicação prática por usuários comuns.

Conclui-se, portanto, que o software livre atua como um alicerce para ambas as vertentes, fomentando a colaboração e acelerando a inovação através de frameworks abertos. O futuro da esteganálise não reside apenas na busca por maior precisão, mas na eficiência e usabilidade. A tendência aponta para o desenvolvimento de arquiteturas, que cada vez mais aumentem a acessibilidade e auditabilidade da segurança da informação.

Referências

Boehm, B. (2014). Stegexpose: A tool for detecting lsb steganography.

Boroumand, M., Chen, M., e Fridrich, J. (2019). Deep residual network for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 14(5):1181–1193.

Chubachi, K. (2020). An ensemble model using cnns on different domains for alaska2 image steganalysis. *IEEE International Workshop on Information Forensics and Security (WIFS)*.

Farooq, N. e Selwal, A. (2023). Image steganalysis using deep learning: a systematic review and open research challenges. *Journal of Ambient Intelligence and Humanized Computing*.

Foundation, F. S. (2024). What is free software? https://www.gnu.org/philosophy/free-sw.html. Acesso em: 27 ago. 2025.

Fridrich, J. (2010). *Steganography in Digital Media 'Principles', Algorithms, and Applications*. Springer.

Fridrich, J., Butora, J., Yousfi, Y., e Khvedchenya, E. (2020). Imagenet pre-trained cnns for jpeg steganalysis. *2020 IEEE International Workshop on Information Forensics and Security (WIFS)*.

Guardian, T. (2025). Apple pulls encrypted icloud storage from uk after government demands back door access. Acesso em: 26 ago. 2025.

Hostalot, D. e Megías, D. (2024). Aletheia: an open-source toolbox for steganalysis. *The Journal of Open Source Software*.

La Croix, J. d. e Ahmad, T. (2023). A convolutional neural network to detect possible hidden data in spatial domain images. *Cybersecurity*.

La Croix, N. J. D., Ahmad, T., e Han, F. (2024). Comprehensive survey on image steganalysis using deep learning. 22.

Poder360 (2025). Europa barrou 41,4 mi de posts via usuários no 1º semestre de 2025. Acesso em: 26 ago. 2025.

Tan, M. e Le, Q. V. (2019a). Efficientnet: Rethinking model scaling for convolutional neural networks.

Tan, M. e Le, Q. V. (2019b). Mixconv: Mixed depthwise convolutional kernels.

Xu, G., Wu, H.-Z., e Shi, Y. Q. (2016). Ensemble of cnns for steganalysis: An empirical study. In *Proceedings of the 8th ACM Workshop on Information Hiding and Multimedia Security*, pages 5–10.

Yousfi, Y. e Fridrich, J. (2020). An intriguing struggle of cnns in jpeg steganalysis and the onehot solution. *IEEE Signal Processing Letters*, pages 830–834.

Yousfi, Y., Fridrich, J., Butora, J., e Tsang, F. C. (2021). Improving efficientnet for jpeg steganalysis. In *Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Secur*, pages 149–157.