1 2  9 0

UNIVERSIDADE D
COIMBRA

Noise simulation for the improvement of printer-proof steganography

Telmo Dias Cunha

Mestrado em Engenharia Eletrotécnica e de Computadores.

Setembro de 2023

Contextualização

- Atualmente, retratos faciais são largamente utilizados como método de verificação de identidade. Desta forma, questões de segurança emergiram, devido a crimes de falsificação de documentos e fraude.

- Tecnologias inovadores como a identificação por radiofrequência (RFID), marca de água e esteganografia permitiram o avanço da segurança digital.

- Esteganografia resistente à impressão permitiu melhorar e criar medidas de segurança que melhoraram a integridade dos documentos de identificação.

- Simulação de ruído presente no ambiente print-scan permite verificar e melhorar a robustez dos documentos de identificação, como também permite ensinar as redes a distinguir ruído, informação e a imagem original.

Motivação

- Como podemos melhorar a segurança dos nossos dados nos documentos de identificação e a sua integridade, na era da tecnologia avançada?
  - ❖ Marca de água que salvaguarda os direitos de autor.
  - ❖ Soluções ponta a ponta de esteganografia que melhoram a segurança de imagens pela ocultação de uma mensagem secreta, através do processo de codificação e descodificação.
  - ❖ Simulação de ruído oferece uma oportunidade para melhorar a robustez das soluções ponta a ponta.
- Indústria e investigadores desenvolveram novos métodos de segurança e ao mesmo tempo surgem novos esquemas de fraude, tornando-se um desafio.

Definição do Problema

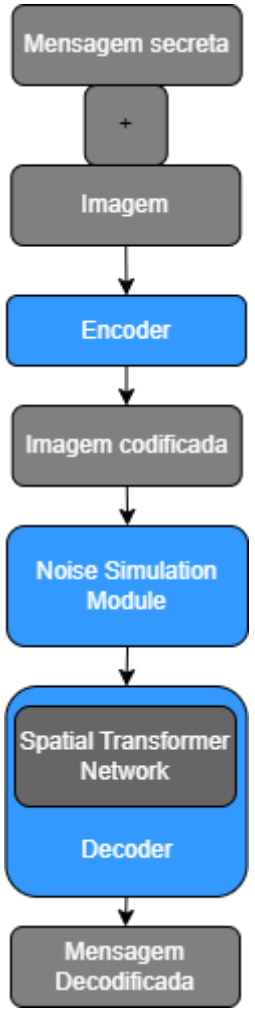
- O módulo de simulação de ruídos de soluções de esteganografia ponta a ponta apresentam pouca robustez durante o processo de impressão e digitalização.
    - ❖ Gama de cores limitada em impressoras.
    - ❖ Distorções introduzidas pela a câmara durante o processo de captura.
    - ❖ Perturbações provenientes do processo de compressão para JPEG.
- Distorções causadas pela incompatibilidade de dispositivos.

Objetivos

- Melhorar e desenvolver a esteganografia resistente à impressão através da simulação de ruído.
- Desafios propostos:
  - ❖ Atingir uma taxa de sucesso perto de 90%.
  - ❖ Aumentar a resistência do modelo enquanto se mantêm a sua performance.

Base (StegaStamp [1])

- Arquitetura do modelo:
  - **Encoder:** Coloca mensagem secreta numa parte específica da rede.
  - **Decoder:** Recupera a mensagem secreta presente na imagem.
  - **Spatial Transformer Network:** Seleciona uma parte específica da imagem, como também retifica a imagem das suas transformações.
  - **Simulação de ruído:** Cria um ambiente realista para a rede durante a fase de treino.



[1] Matthew Tancik, Bem Mildenhall, and Ren Ng. StegaStamp: Invisible hyperlinks in physical photographs. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 2117-2126, 2020.

Simulação de ruído

- **Planckian Jitter:** Centra-se nas variações de iluminação para uma precisão mais exata da tonalidade e saturação.
- **Poisson noise:** Incerteza na medição da luz devido à natureza quantizada da luz e à independência da deteção de fotões.



Simulação de ruído (cont.)

- **Dark noise:** Incerteza na dark current devido a flutuações estatísticas nos eletrões gerados termicamente.

- **Speckle noise:** Ruído granular que degrada a qualidade da imagem devido à interferência entre wavefronts em sistemas de imagem.

Simulação de ruído (cont.)

- **Misregistration noise:** Simula o ruído que resulta do desalinhamento dos canais da imagem.
- **Motion Blur:** Efeito óptico causado pelo movimento do objeto durante a abertura do obturador da câmara.

Simulação de ruído (cont.)

- **Posterization:** Simplifica a tonalidade de uma imagem para áreas planas distintas, reduzindo o número de tons.

- **Plasma Brigthness:** Refere-se a uma variação aleatória do brilho que afeta uma imagem, resultando numa imagem com padrões aleatórios.

Dual Contrastive Loss

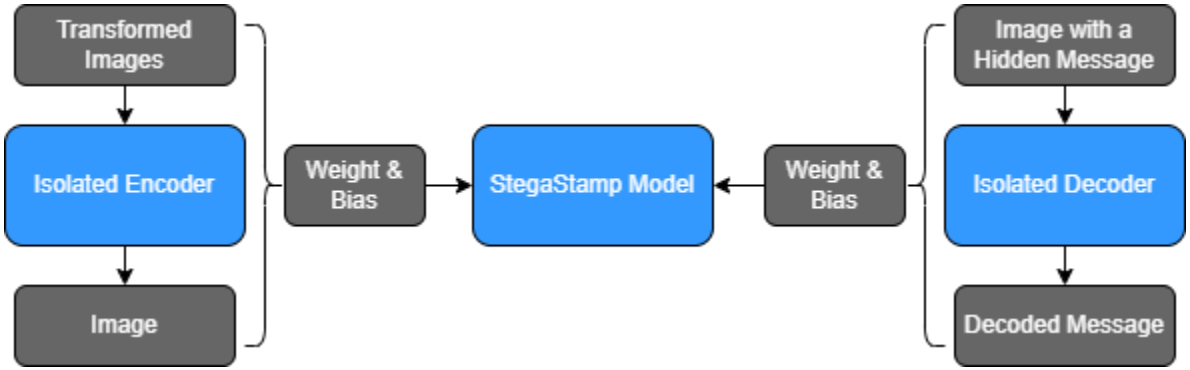
- Tem como objetivo mapear amostras semelhantes mais próximas umas das outras e amostras diferentes mais afastadas no espaço de incorporação.
- Constituído por dois componentes chaves: pares positivos e pares negativos.

Data augmentation

- **Métodos Convencionais:** Aplicação de combinação de transformações.
- **Neural Style Transfer:** Cria uma nova imagem através da extração do conteúdo e estilo de imagens diferentes a partir de uma rede neuronal profunda pré-treinada.

Self-Supervising learning (SSL)

- Permite que o modelo aprenda a representação dos dados sem labels manuais.
- **Encoder:** Constituído por uma rede U-net, permite a realização de um problema de restauração da informação.
- **Decoder:** Abordar-se uma SSL de forma a melhorar o processo de descodificação a partir de um dataset com mensagens ocultas.

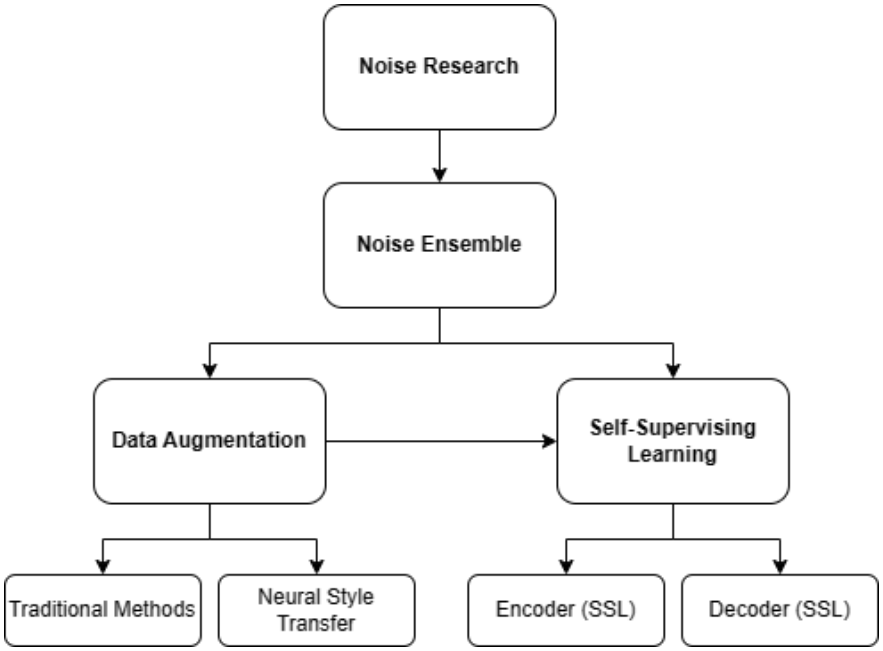


Métricas

- **Structural Similarity Index (SSIM):** Avalia a informação estrutural, a luminância e as semelhanças de contraste entre imagem de referência e a imagem distorcida.
- **Peak Signal-to-Noise Ratio (PSNR):** Avalia a qualidade de uma imagem comparando os seus pixéis com a imagem de referência para identificar distorções.
- **Taxa de descodifação:** Quantifica o número de imagem codificadas que foram descodificadas com sucesso.

Plano de Simulação

- **Noise research:** Investiga e avalia o impacto do ruído no modelo, durante o processo print-scan.
- **Noise Ensemble:** Avalia o modelo sob a influência de vários ruídos.
- **Data Augmentation:** Melhora a performance do modelo consoante o aumento gradual de amostras no dataset.
- **Self-Supervising Learning:** Melhora a performance do modelo usando Self-Supervising Learning em partes específicas do modelo e avalia a sua eficácia.



Datasets

- **MIRFLICKR dataset:** Utilizado para treinar o modelo StegaStamp.
- **JMiPOD dataset:** Utilizado com o propósito de criar um problema para o decoder de forma a permitir o uso de SSL.

Condições de Teste

- Testado em ambiente digital.
- Tamanho das imagens: 400*400.
- Qualidade da imagem: 70-300 dpi (dots per inch).
- Qualidade da imagem(codificada): 96 dpi.

Resultados Base

- Base 1 e Base 2 representam o teste base com e sem a utilização de Spacial Transformation Network, respetivamente.

Test	Epochs	Decoding rate
Base 1	140,000	70.3%
Base 2	140,000	80.4%

- A incorporação de STN no modelo aumenta significativamente a complexidade do modelo.

Ruído individual

	Noise			
	Posterization	Planckian Jitter	Poisson	Misregistration
Decoding rate	75,1%	78,7%	79,9%	84,6%
ΔBase 1	↑ 4,8 pp	↑ 8,4 pp	↑ 9,6 pp	↑ 14,3 pp
SSIM	0,71	0,72	0,68	0,69
PNSR	52,0	52,3	51,6	51,8
	Noise			
	Dark	Plasma Brightness	Motion Blur	Speckle
Decoding rate	77,1%	74,8%	78,7%	80,0%
ΔBase 1	↑ 6,8 pp	↑ 4,5 pp	↑ 8,4 pp	↑ 9,7 pp
SSIM	0,69	0,68	0,70	0,69
PNSR	63,7	63,5	64,1	63,9

- Impacto de cada ruído no modelo varia de acordo com a sua influência na imagem, como também com a sua natureza.
- O modelo encontra-se no seu estado de melhor performance.

Grupo de ruídos

- Grupo 1: Posterização, Planckian Jitter e Poisson noise.
- Grupo 2: Grupo 1 com a adição de Misregistration noise.
- Grupo 3: Grupo 2 com a adição de Motion Blur.

Noise group	Epochs	Decoding rate	ΔBase 1	SSIM	PSNR
Group 1	140.000	81,4%	↑ 11,1 pp	0,71	64,4
	180.000	78,3%	↑ 8,0 pp	0,69	51,7
	140.000	83.6%	↑ 12,3 pp	0,69	57,9
Group 2	180.000	81,3 %	↑ 11,0 pp	0,69	63,9
Group 3	160.000	79,1%	↑ 8,8 pp	0,71	51,8
	180.000	74,6%	↑ 4,3 pp	0,70	51,9

- O aumento do número de ruídos é acompanhado por uma redução do desempenho global do modelo.

- Desenvolvimento de um algoritmo complementar com o objetivo de equilibrar as influencias de cada fonte de ruído e diminuir a influência do método tentativa e erro.

Data Augmentation

Noise combination	Epochs	Decoding rate	ΔBase 1	SSIM	PSNR
Group 1	140.000	83,0 %	↑ 12,7 pp	0,73	52,4
Group 2	180.000	82,3 %	↑ 12,0 pp	0,69	51,9
Group 3	160.000	81,7 %	↑ 11,4 pp	0,70	52,0

Resultados obtidos com Grupo 1

Dataset size	Decoding rate	ΔBase 1	SSIM	PSNR
100.000	83,7 %	↑ 13,4 pp	0,71	52,1
200.000	80,0 %	↑ 9,7 pp	0,73	52,4

- A duplicação de amostras presente no dataset produz um crescimento modesto de 2pp.
- O aumento de amostras no dataset para 100.000 e 200.000 revela resultados poucos satisfatórios.

Data Augmentation (cont.)

- O uso de Neural Style Transfer provou ser inadequada, uma vez que as imagens obtidas apresentavam uma qualidade inferior ou estruturas inadequadas para o problema em questão.

Self-Supervising learning

- Resultado obtidos com o Grupo 1 com a utilização de pré-treino no encoder.

Decoding rate	ΔBase 1	SSIM	PSNR
82,7 %	↑ 12,4 pp	0,71	58,3

- **Encoder:**
  - ❖ Apesar do resultado ser satisfatório, demonstra necessidade de ajustamento dos parâmetros do modelo.
- **Decoder:**
  - ❖ Não foi possível obter resultados com a implementação de SSL.
  - ❖ O dataset escolhido apresentava mensagens ocultas com diferentes tamanhos.
  - ❖ Desempenho inadequado do processo de recuperação de imagens à priori.

Análise global

- Resultados obtidos revelam a eficácia do plano de simulação definido.
- A escolha de um dataset diferente com mais amostras pode resultar em melhores valores.
- Desenvolver uma abordagem que diminua a influência de tentativa e erro.
- Em comparação com o estado da arte, Self-Supervising learning constitui uma nova abordagem.
- A realização do processo de esteganografia em ambiente físico permite uma melhor avaliação da eficácia das hipóteses desenvolvidas.

Conclusão

- Simulação de ruído provou ser um método robusto e eficiente para melhorar a esteganografia à prova de impressão.
- Ilustra o desafio de aumentar a robustez do sistema sem comprometer o desempenho do modelo.
- O estudo desenvolvido contribuiu para a melhoria da esteganografia resistente à impressão através do desenvolvimento do módulo simulação de ruído.

Agradecimentos

Professor Ph.D Nuno Gonçalves

Ph.D. Luiz Schirmer

VIS Team do Instituto de Sistemas e Robótica

Departamento de Engenharia Eletrotécnica e de Computadores

Universidade de Coimbra