Uso de Autoencoders para Análise de Imagens no Experimento CYGNO

Antonio Marchese

Engenharia Elétrica - Sistemas Eletrônicos

Universidade Federal de Juiz de Fora (UFJF)

14 de julho de 2026

Resumo

O experimento CYGNO atua na detecção direta de matéria escura através de uma Câmara de Projeção Temporal (TPC) com leitura óptica baseada em sensores sCMOS de alta resolução. Embora ofereça excelente sensibilidade direcional, a tecnologia gera um fluxo massivo de dados em que a imensa maioria das imagens contém estritamente ruído eletrônico e óptico, o chamado pedestal. Para evitar o armazenamento inviável de dados irrelevantes, é fundamental o desenvolvimento de algoritmos de *trigger* online rápidos para filtrar o sinal de fundo e isolar Regiões de Interesse (ROIs) em tempo real. Este trabalho propõe e valida um filtro computacional baseado em aprendizado de máquina auto-supervisionado, utilizando uma arquitetura de *Denoising Autoencoder* (DAE) integralmente retreinada a partir do zero. O modelo extrai as características estocásticas do ruído ao ser alimentado com imagens corrompidas por anomalias sintéticas, sendo forçado a reconstruir apenas o pedestal limpo através de uma função de perda que pondera o Erro Quadrático Médio (MSE) e o Índice de Similaridade Estrutural (SSIM). Na fase de inferência, a operação subtrativa revela as assinaturas físicas reais destacadas no mapa de resíduos. Testes realizados com simulações Monte Carlo (GEANT4) demonstraram que a rede preserva a integridade morfológica dos rastros e assegura um Ganho de Relação Sinal-Ruído (SNR_{gain}) superior a 1, mesmo atuando no limite da sensibilidade do detetor em regimes de sub-keV (0.25 keV). Além disso, a etapa de supressão espacial (*zero-suppression*) reteve tipicamente menos de 10% da área da matriz em quadros com eventos válidos, convergindo o armazenamento de alarmes falsos a zero em amostras de ruído puro. Os resultados comprovam a alta capacidade discriminativa e de compressão do método, delineando como trabalhos futuros a análise sistemática do tempo de execução em hardware e a validação física utilizando sinais reais coletados pelo protótipo LIME.

Sumário

1 Introdução **3**

2 Fundamentação Teórica **5**
2.1 O Experimento CYGNO 5
2.2 O Protótipo LIME 6
2.3 Redes Neurais Convolucionais (CNNs) 8
2.4 *Autoencoders* e Aprendizado Auto-Supervisionado 9
2.5 Deteção de Anomalias e Métricas de Reconstrução 10

3 Metodologia **12**
3.1 Pré-processamento dos Dados 12
3.2 Arquitetura do *Autoencoder* Convolucional 13
3.3 Treinamento e Pós-processamento 14
3.3.1 Modelagem Matemática das Anomalias Sintéticas 14
3.3.2 Fluxo de Propagação e Penalização Espacial 16
3.3.3 Inferência e Extração da Região de Interesse (ROI) 16

4 Resultados e Discussões **18**
4.1 Descrição do Conjunto de Dados 18
4.2 Avaliação Morfológica do Sinal Extraído via SSIM Local 19
4.3 Análise do Ganho de Relação Sinal-Ruído (SNR) 20
4.4 Análise da Região de Interesse (ROI) e Compressão de Dados 22

5 Conclusão e Trabalhos Futuros **23**

1 Introdução

A deteção direta de matéria escura é um dos grandes desafios da física de partículas contemporânea. Como essa forma de matéria é invisível e raramente interage com o nosso mundo, os cientistas precisam construir detectores extremamente sensíveis para realizar a sua detecção. O experimento CYGNO propõe uma abordagem inovadora para essa busca: ele funciona, essencialmente, como uma câmera tridimensional ultrassensível apontada para um volume de gás. Quando uma partícula esbarra nas moléculas desse gás, ela cria uma trilha de energia que é amplificada e transformada em luz (luminescência). Para registrar esse rastro, o experimento utiliza um sistema duplo: câmeras científicas de altíssima resolução (sCMOS) tiram "fotos"da trilha em duas dimensões (eixos X e Y), enquanto sensores, chamados Tubos Fotomultiplicadores (PMTs), medem a fração de segundo que a luz leva para chegar até as lentes, o que permite calcular a profundidade (eixo Z) e reconstruir a trajetória da partícula em 3D (Amaro, 2026).

Apesar de a alta granularidade do sistema de leitura óptica proporcionar uma excelente sensibilidade, ela introduz um desafio de aquisição e armazenamento de dados. O experimento produz imagens na escala de megapixels que, sem um mecanismo de seleção *online*, geram um volume de dados massivo (Amaro, 2026). Na operação diária, a expectativa é de milhares de quadros, cuja maioria contém estritamente ruído eletrônico e óptico, além de artefatos do sensor, sem nenhum evento físico relevante (Pains, 2024). Armazenar integralmente todo esse fluxo para análise *offline* é inviável, tornando indispensável o desenvolvimento de algoritmos de *trigger* rápidos, capazes de filtrar o sinal de fundo e extrair Regiões de Interesse (ROIs) em tempo real (Amaro, 2026).

Estudos anteriores no contexto do experimento CYGNO buscaram solucionar esse gargalo computacional. Na própria Universidade Federal de Juiz de Fora (UFJF), investigaram-se algoritmos de *trigger* focados tanto em técnicas clássicas de filtragem espacial, incluindo métodos determinísticos, como filtros gaussianos e estatísticos, como o filtro casado, quanto em classificação binária supervisionada baseada em Redes Neurais Convolucionais (CNNs) (Pains, 2024). Embora abordagens baseadas em classificação supervisionada atinjam altas taxas de detecção para sinais conhecidos e mantenham alarmes falsos consideravelmente baixos, elas exigem rotulações rígidas e dependem excessivamente de simulações, o que pode comprometer a generalização em relação a sinais anômalos não contemplados durante o treinamento (Pains, 2024).

Diante dessas limitações, os métodos fundamentados em aprendizado de máquina autosupervisionado (*Self-Supervised Learning*) surgem como uma alternativa robusta e genérica. A estratégia central deste Trabalho de Conclusão de Curso (TCC) baseia-se na arquitetura originalmente proposta no âmbito da colaboração CYGNO (Amaro, 2026). O objetivo principal deste estudo consiste em reproduzir de forma independente essa topologia computacional e submetê-la a testes sob condições de baixa energia (regimes de sub-keV), mapeando de forma sistemática os limites de sensibilidade do filtro e analisando potenciais janelas de otimização e melhorias estruturais na supressão do ruído.

Diferentemente da mera importação dos pesos e da estrutura pré-treinada utilizada no artigo original, este trabalho realizou o re-treinamento integral da rede a partir do zero. Essa abordagem permite avaliar a estabilidade de convergência do algoritmo sob novas distribuições estocásticas de ruído, além de consolidar uma implementação própria da *pipeline* de processamento de tensores de dados do protótipo LIME (Marchese, 2026). Sob o paradigma de *Denoising Autoencoder* (DAE), o modelo é alimentado com imagens corrompidas por injeções de anomalias sintéticas lineares e pontuais, mas é forçado a re-

construir estritamente o pedestal limpo (Vincent *et al.*, 2008; Amaro, 2026). Na fase de inferência, o algoritmo pondera funções de perda como o Erro Quadrático Médio (MSE) e o Índice de Similaridade Estrutural (SSIM) para omitir as trilhas de ionização reais. O propósito final é validar uma ferramenta computacional leve, eficiente para inferência em GPU e agnóstica em relação ao modelo teórico do sinal, fornecendo uma base escalável para os requisitos de *trigger online* nas futuras etapas operacionais do experimento CYGNO.

2 Fundamentação Teórica

Este capítulo apresenta os conceitos físicos e computacionais que embasam a metodologia proposta. Inicialmente, descreve-se o princípio de funcionamento do experimento CYGNO e a origem dos dados visuais. Em seguida, abordam-se os fundamentos das Redes Neurais Convolucionais e dos *Autoencoders*, culminando nas métricas utilizadas para a deteção de anomalias em imagens.

2.1 O Experimento CYGNO

O experimento CYGNO representa um esforço internacional dedicado à deteção direta e direcional de matéria escura, focado primariamente na procura por partículas massivas de interação fraca (WIMPs) com massas na escala de baixos GeV (Ereditato, 2022). Para atingir a sensibilidade direcional necessária, a colaboração propõe a utilização de um detetor gasoso do tipo *Time Projection Chamber* (TPC), operando à pressão atmosférica com uma mistura de gás Hélio e Tetrafluoreto de Carbono (He:CF_4) (Ereditato, 2022; Amaro, 2026). As interações esperadas no gás detector devem depositar energias da ordem de poucos keV, produzindo um número reduzido de elétrons primários de ionização (1 keV depositado no gás hélio produz cerca de 25 elétrons), posteriormente amplificados pelo sistema de multiplicação eletrônica baseado em detectores do tipo GEM (Gas Electron Multipliers) cuja avalanche eletrônica induz a emissão de fótons no gás, produzindo sinais ópticos de baixa intensidade.

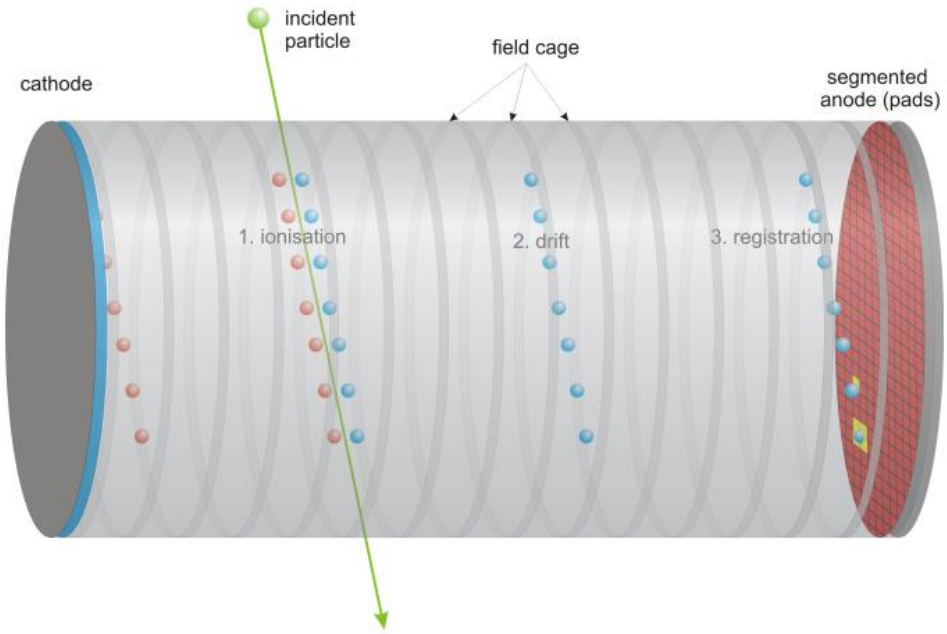


Figura 1: Ilustração do princípio de funcionamento de uma *Time Projection Chamber* (TPC), demonstrando a trajetória da partícula incidente, a ionização e o campo de deriva.

Fonte: (LCTPC Collaboration, 2026)

O princípio de detecção baseia-se na captura da luz emitida durante o processo de multiplicação de elétrons. Quando uma partícula incidente interage com o gás na TPC,

esta ioniza as moléculas ao longo da sua trajetória, liberando elétrons primários (Ereditato, 2022). Um campo elétrico uniforme, aplicado através do volume do detetor, faz com que estes elétrons derivem em direção à região de amplificação, composta por uma estrutura de múltiplos *Gas Electron Multipliers* (GEMs) (Ereditato, 2022; Pains, 2024). Nos buracos dos GEMs, o elevado campo elétrico induz cascatas de ionização, multiplicando o número de elétrons e gerando fotóns de cintilação através de um processo de eletroluminescência (Ereditato, 2022).

A inovação central do CYGNO reside no seu sistema de leitura (*readout*) ótica. A luz produzida nos GEMs é captada por câmaras de tecnologia *scientific* CMOS (sCMOS), acopladas a objetivas de grande abertura (Ereditato, 2022; Pains, 2024). Esta abordagem ótica permite obter imagens 2D com uma granularidade espacial extremamente elevada da projeção dos rastros das partículas no plano do GEM (Amaro, 2026). Em conjunto com a temporização do sinal obtida por fotomultiplicadores (PMTs), o sistema permite uma reconstrução tridimensional precisa da energia depositada e da direção do recuo (Ereditato, 2022).

Contudo, a sensibilidade extrema dos sensores sCMOS traz consigo desafios significativos ao processamento e ao sistema de aquisição de dados (DAQ). O sensor registra não apenas o sinal luminoso proveniente das interações físicas, mas também contribuições associadas ao ruído eletrônico intrínseco, à corrente de escuro e a flutuações espaciais do dispositivo, compondo o chamado pedestal ou imagem de fundo estacionária (Pains, 2024). Como detalhado por Amaro et al. (Amaro, 2026), o futuro demonstrador CYGNO-04 — um TPC de 0,4 m^3 equipado com dois planos ópticos de 50×80 cm^2 — registrará imagens com resolução de 4096×2304 pixels. Cada quadro isolado corresponde a aproximadamente 18,9 MB. Operando seis câmeras simultâneas a uma taxa de captura de cerca de 3 Hz, o experimento produzirá aproximadamente 18 imagens por segundo, resultando num fluxo contínuo de dados que excede 340 MB/s. Este volume massivo torna-se rapidamente incomportável para armazenamento integral e análise *offline*, tornando indispensável o desenvolvimento de processos de filtragem espacial e algoritmos de *trigger* eficientes.

A imensa maioria destas aquisições representa unicamente o pedestal do sensor, sem a presença de rastros de partículas relevantes (Pains, 2024). Consequentemente, para viabilizar a gravação seletiva e a análise posterior (*offline*), torna-se imperativa a implementação de um sistema de *trigger online* robusto (Amaro, 2026; Pains, 2024). É neste contexto de isolamento de sinais raros imersos num fundo ruidoso de alta resolução que a aplicação de modelos de reconstrução espacial e deteção de anomalias – como os *Autoencoders* – se justifica como uma solução arquiteturalmente viável.

2.2 O Protótipo LIME

A colaboração CYGNO propõe o desenvolvimento de um detetor com elevada resolução espacial e de energia, otimizado para a procura direcional de Matéria Escura, focando-se na deteção de recuos nucleares de baixa energia gerados pela dispersão elástica de WIMPs (*Weakly Interacting Massive Particles*) (Ereditato, 2022). O protótipo faz uso de uma Câmara de Projeção Temporal (TPC) operada à pressão atmosférica, utilizando uma mistura gasosa de Hélio e Tetrafluorometano (He/CF_4), tipicamente na proporção 60/40 (Amaro, 2026; Pains, 2024). Esta escolha garante a sensibilidade tanto a interações dependentes de spin (através do Flúor) quanto independentes de spin (Hélio), além de o CF_4 atuar como um excelente cintilador no espectro visível.

O detetor LIME delimita um volume ativo de 50 litros contido num vaso de acrílico

(PMMA) transparente. O seu campo elétrico de deriva (*drift*) é estabelecido por uma gaiola de campo composta por anéis de cobre, estendendo-se por uma profundidade máxima de 50 cm entre o cátodo e a região de amplificação.

A amplificação do sinal de ionização ocorre de forma mecanicamente isolada do volume alvo, utilizando uma cascata de três camadas de *Gas Electron Multipliers* (Triplo-GEM) com uma área sensível de 330×330 mm^2. Quando elétrons primários derivados do evento físico atingem os orifícios dos GEMs, eles sofrem um processo de avalanche eletrônica. Devido à presença do gás CF_4, esta multiplicação de carga é acompanhada pela emissão de um forte sinal de luminescência, cuja intensidade é estritamente proporcional à energia cinética depositada pela partícula inicial (Pains, 2024).

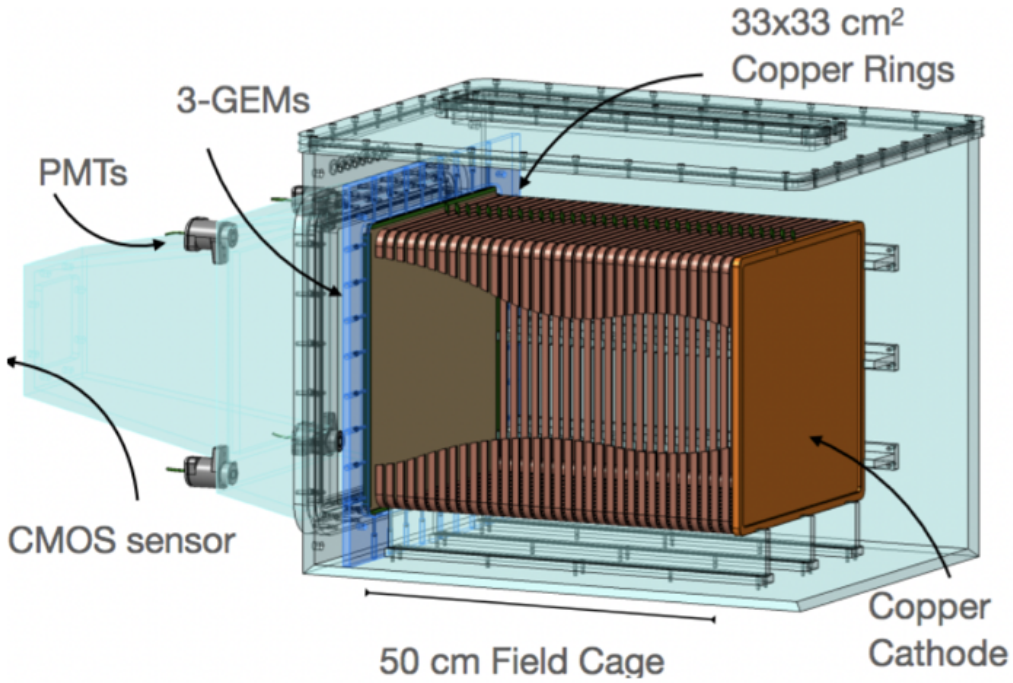


Figura 2: Renderização esquemática do protótipo LIME.

Fonte: Adaptado de Baracchini et al. (Baracchini *et al.*, 2023).

A grande inovação arquitetural do LIME concentra-se no seu sistema de leitura ótica duplo, localizado externamente ao volume de gás para evitar a contaminação cruzada por radioatividade de materiais:

- **Câmera sCMOS:** Um sensor científico de alta granularidade (modelo ORCA Quest ou Flash), focado opticamente na face do último GEM. É responsável por integrar a luz cintilante e gerar projeções bidimensionais (plano X-Y) de alta resolução. A câmera produz imagens de 2304×2304 pixels e registra a morfologia do rastro.
- **Tubos Fotomultiplicadores (PMTs):** Um arranjo de quatro PMTs de alta eficiência acoplados à janela ótica do detetor, operando com uma resposta temporal na ordem de nanossegundos. Conforme demonstrado de forma pioneira pela colaboração CYGNO (Amaro *et al.*, 2023), enquanto a sCMOS projeta o rastro no plano transversal, os PMTs amostram o perfil temporal da emissão de luz contínua.

Conhecendo-se a velocidade de deriva (*drift velocity*) dos elétrons no gás He/CF_4, a diferença nos tempos de chegada dos fótons permite calcular a coordenada relativa da terceira dimensão (ΔZ). Além disso, a análise do alargamento temporal do sinal luminoso induzido pela difusão longitudinal viabiliza a inferência da coordenada absoluta Z da interação original, convertendo o detetor num verdadeiro rastreador 3D (Amaro *et al.*, 2023).

A matriz produzida pela câmera sCMOS do protótipo LIME atua como matéria-prima para o presente trabalho. As imagens brutas geradas pela leitura fotográfica, sobrecarregadas pelo ruído eletrônico inerente ao sensor e por artefactos luminosos, impõem o desafio fundamental de separação de sinal e fundo que o algoritmo de *trigger* baseado em *Autoencoders* visa solucionar em tempo real.

2.3 Redes Neurais Convolucionais (CNNs)

O processamento de imagens de alta resolução, como as geradas pelo experimento CYGNO, exige modelos computacionais capazes de extrair padrões espaciais complexos de forma eficiente. As Redes Neurais Convolucionais (CNNs - *Convolutional Neural Networks*) consolidaram-se como a principal arquitetura para tarefas de visão computacional, superando as redes densamente conectadas tradicionais ao preservar a topologia espacial dos dados e reduzir drasticamente o número de parâmetros treináveis (Pains, 2024).

A operação fundamental de uma CNN é a convolução. Neste processo, um filtro (ou *kernel*) bidimensional desliza sobre a imagem de entrada, realizando produtos escalares locais. Essa operação é frequentemente implementada como uma correlação cruzada, permitindo que a rede aprenda a identificar características morfológicas progressivas, como perturbações básicas, bordas (nas primeiras camadas) e até rastros de ionização (em camadas mais profundas). Matematicamente, a operação de convolução bidimensional de uma imagem I por um *kernel* K de dimensões $M \times N$ pode ser expressa como:

$$S(i,j) = (I * K)(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(i-m, j-n) K(m,n) \tag{1}$$

onde $S(i,j)$ representa o valor do mapa de características (*feature map*) gerado na posição espacial (i,j) após a aplicação do filtro.

A fim de garantir uma representação fidedigna do ruído de fundo (pedestal) do sensor sCMOS, é essencial que a rede seja capaz de modelar relações estocásticas e não lineares. Para isso, funções de ativação são aplicadas ao resultado das convoluções. Funções não lineares, como a ReLU (*Rectified Linear Unit*) e suas variantes, são amplamente utilizadas, pois aceleram a convergência e mitigam o problema do desaparecimento do gradiente durante o treinamento.

Além das camadas convolucionais, arquiteturas baseadas em imagens frequentemente empregam operações de redução de dimensionalidade espacial (*downsampling*). Isso é feito através de camadas de *Pooling* (agrupamento) ou de convoluções com passos (*strides*) maiores que um. Essa compressão do espaço de características é crucial não apenas para diminuir o custo computacional em tempo de inferência, mas também para forçar a rede a extrair as informações mais representativas dos dados. Esse exato princípio de compressão e extração de variáveis latentes forma a base conceitual para os modelos de reconstrução que fundamentam este trabalho.

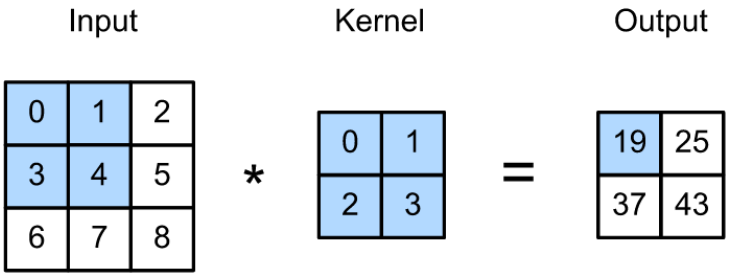


Figura 3: Ilustração da operação de convolução bidimensional utilizada em CNNs (implementada na prática como correlação cruzada). As regiões em azul demonstram o primeiro cálculo do produto escalar entre o tensor de entrada e o kernel.

Fonte: Adaptado de Zhang et al. (Zhang *et al.*, 2023).

2.4 *Autoencoders* e Aprendizado Auto-Supervisionado

No contexto do experimento CYGNO, a escassez de eventos reais de interação de matéria escura e a complexidade em mapear manualmente grandes volumes de imagens contendo diferentes variações de traços tornam as abordagens de aprendizado supervisionado convencional limitadas (Pains, 2024). Como alternativa robusta, o aprendizado auto-supervisionado (*Self-Supervised Learning* - SSL) surge como um paradigma que foca em extrair representações profundas diretamente da estrutura dos dados, sem a necessidade de rótulos ou anotações prévias fornecidas por especialistas humanos.

É sob este arranjo que os *Autoencoders* operam neste trabalho. Ao projetar uma tarefa de pretexto baseada na injeção estocástica de anomalias sintéticas e exigir matematicamente que a rede aprenda a restaurar o rastro ocultado para alcançar o alvo limpo, o sistema constrói as suas próprias restrições de supervisão de forma endógena (Vincent *et al.*, 2008). Essa estratégia auto-supervisionada permite que o modelo aprenda de maneira autônoma a morfologia estatística do ruído de fundo, garantindo alta sensibilidade a desvios anômalos sem depender de um banco de dados real previamente catalogado.

Conforme definido por Bank, Koenigstein e Giryes (Bank; Koenigstein; Giryes, 2020), um *Autoencoder* é um tipo específico de rede neural concebida fundamentalmente para codificar uma entrada numa representação comprimida e significativa e, em seguida, decodificá-la, de tal forma que a saída reconstruída seja o mais semelhante possível à original. Estruturalmente, um *Autoencoder* clássico é composto por três componentes principais:

1. ***Encoder* (Codificador):** Uma rede neural, parametrizada por θ, que recebe uma imagem de entrada $x \in \mathbb{R}^N$ e a mapeia para uma representação compacta $z \in \mathbb{R}^M$, onde $M \ll N$. Esta transformação pode ser descrita como $z = f_\theta(x)$. Tratando-se de imagens, tipicamente o *encoder* é composto por camadas convolucionais sequenciais, a fim de reduzir a dimensionalidade espacial enquanto aumenta a profundidade dos canais.

2. **Espaço Latente (*Bottleneck*):** O vetor z resultante é conhecido como espaço latente. Ele atua como um gargalo de informação (*bottleneck*), forçando a rede a descartar a reter apenas as características fundamentais (variáveis latentes) necessárias para descrever os dados de forma abstrata.

3. ***Decoder* (Descodificador):** Uma segunda rede, parametrizada por ϕ, que recebe o vetor latente z e tenta reconstruir a imagem original, gerando uma saída $x' \in \mathbb{R}^N$. Esta operação é expressa como $x' = g_\phi(z) = g_\phi(f_\theta(x))$. Nas CNNs, este processo é realizado através de camadas de deconvolução (ou convoluções transpostas) e operações de *upsampling*.

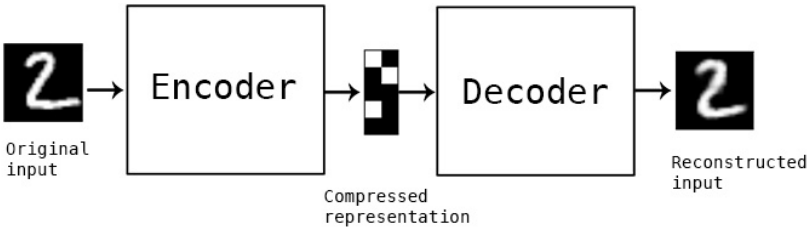


Figura 4: Arquitetura geral de um *Autoencoder*, ilustrando o fluxo de compressão da imagem original pelo *Encoder* até o espaço latente e a subsequente reconstrução pelo *Decoder*.

Fonte: Adaptado de Chollet (Chollet, 2016).

O treinamento de um *Autoencoder* consiste em minimizar uma função de perda (*Loss Function*) que quantifica a divergência entre a entrada original x e a reconstrução x'.

A aplicação desta arquitetura para a deteção de anomalias no CYGNO baseia-se no paradigma de *Denoising Autoencoders* (DAE) (Vincent *et al.*, 2008). Durante a fase de treino, o modelo é alimentado com imagens de pedestal corrompidas pela injeção estocástica de anomalias sintéticas (rastros lineares e aglomerados gaussianos) (Marchese, 2026). Contudo, o alvo (*target*) de reconstrução exigido pela função de perda permanece estritamente o fundo limpo do sensor sCMOS (ruído eletrónico e ótico estacionário) (Amaro, 2026). Ao ser forçado a suprimir as estruturas artificiais inseridas, o modelo aprende a mapear o espaço latente exclusivamente para a topologia estatística específica do ruído, tornando-se cego a assinaturas morfológicas de sinal.

Durante a fase de inferência (*triggering*), ao receber uma imagem real contendo o rastro de ionização de uma partícula física, o *Autoencoder* atua como um filtro seletivo de supressão. O modelo estima o pedestal subjacente e elimina a componente anômala da imagem reconstruída x'. Como o rastro físico é ativamente omitido pelo *decoder*, a operação subtrativa absoluta pixel a pixel entre a entrada (x) e a saída (x') resulta num erro de reconstrução localmente elevado exatamente nas coordenadas geométricas do evento.

2.5 Deteção de Anomalias e Métricas de Reconstrução

O sucesso do sistema de *trigger* baseado em *Autoencoders* depende diretamente da escolha das métricas utilizadas para avaliar o erro de reconstrução. Uma função de perda inadequada pode fazer com que a rede ignore detalhes estruturais finos dos rastros de ionização, comprometendo a precisão na extração das Regiões de Interesse (ROIs). Para contornar este problema, utiliza-se uma abordagem que combina critérios estatísticos ponto a ponto e métricas de similaridade estrutural.

A métrica mais tradicional para a avaliação de reconstrução de imagens é o Erro Quadrático Médio (MSE - *Mean Squared Error*). O MSE calcula a média das diferenças elevadas ao quadrado entre os pixels da imagem original x e os da imagem reconstruída x'. Para duas imagens de dimensões $W \times H$, o MSE é definido por:

$$\mathrm{MSE}(x, x') = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} (x(i,j) - x'(i,j))^2 \tag{2}$$

Embora o MSE seja computacionalmente eficiente e possua propriedades matemáticas favoráveis para a otimização, ele apresenta limitações na preservação de estruturas de alta frequência espacial. Isso acontece pois ele assume independência estatística entre pixels vizinhos, tendendo a gerar reconstruções excessivamente suavizadas (desfocadas), o que pode mascarar os limites precisos de um rastro de partícula ténue imerso no ruído.

Para superar esta limitação, incorpora-se o Índice de Similaridade Estrutural (SSIM - *Structural Similarity Index Measure*) (Wang *et al.*, 2004). O SSIM avalia a similaridade entre duas imagens focando em três componentes independentes do sistema visual humano: luminância (l), contraste (c) e estrutura (s). Dados dois blocos locais de imagem, as componentes são calculadas com base nas médias (μ), variâncias (σ^2) e covariância ($\sigma_{xx'}$):

$$l(x, x') = \frac{2\mu_x\mu_{x'} + c_1}{\mu_x^2 + \mu_{x'}^2 + c_1}, \quad c(x, x') = \frac{2\sigma_x\sigma_{x'} + c_2}{\sigma_x^2 + \sigma_{x'}^2 + c_2}, \quad s(x, x') = \frac{\sigma_{xx'} + c_3}{\sigma_x\sigma_{x'} + c_3} \tag{3}$$

onde c_1, c_2 e c_3 são constantes de estabilização para evitar divisões por zero. Combinando estas três funções com pesos iguais, a fórmula simplificada do SSIM assume a forma:

$$\mathrm{SSIM}(x, x') = \frac{(2\mu_x\mu_{x'} + c_1)(2\sigma_{xx'} + c_2)}{(\mu_x^2 + \mu_{x'}^2 + c_1)(\sigma_x^2 + \sigma_{x'}^2 + c_2)} \tag{4}$$

O valor do SSIM varia no intervalo $[-1, 1]$, onde a unidade indica uma reconstrução perfeita. Durante o treino do modelo, a função de perda pode ser otimizada utilizando uma combinação ponderada, como a perda por similaridade estrutural ($1 - \mathrm{SSIM}$) associada ao MSE, forçando a rede a reter os detalhes texturais do pedestal.

Uma vez treinado o modelo, a deteção da anomalia ocorre através da análise do mapa de resíduos gerado pela diferença absoluta (pixel a pixel) entre a imagem original e a imagem de saída da rede. Para validar a eficácia do algoritmo em destacar o rastro da partícula frente ao fundo ruidoso original, utiliza-se a métrica de Relação Sinal-Ruído (SNR). Neste contexto, a SNR de uma matriz é classicamente definida pela razão entre a intensidade média dos pixels pertencentes à região geométrica do sinal (μ_{sinal}) e o desvio padrão dos pixels pertencentes exclusivamente à região de fundo estacionário (σ_{fundo}):

$$\mathrm{SNR} = \frac{\mu_{\mathrm{sinal}}}{\sigma_{\mathrm{fundo}}} \tag{5}$$

A partir desta definição, avalia-se o Ganho de Relação Sinal-Ruído ($\mathrm{SNR}_{\mathrm{gain}}$) proporcionado pelo algoritmo (Amaro, 2026). Este ganho é definido pela razão direta entre a SNR medida no mapa de resíduos após o processamento pelo *Autoencoder* ($\mathrm{SNR}_{\mathrm{final}}$) e a SNR intrínseca contida na imagem de entrada ($\mathrm{SNR}_{\mathrm{inicial}}$):

$$\mathrm{SNR}_{\mathrm{gain}} = \frac{\mathrm{SNR}_{\mathrm{final}}}{\mathrm{SNR}_{\mathrm{inicial}}} \tag{6}$$

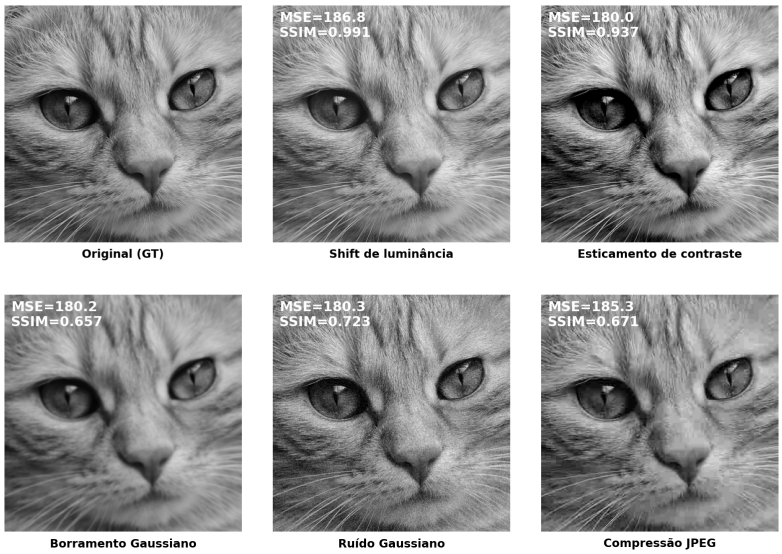


Figura 5: Comparação empírica entre os índices MSE e SSIM para diferentes tipos de artefatos e degradações em relação à imagem de referência.

Fonte: O autor.

Um $\mathrm{SNR}_{\mathrm{gain}} > 1$ indica que o algoritmo reduziu a variância do ruído de fundo sem atenuar severamente a intensidade do sinal da partícula anómala, validando quantitativamente a capacidade do modelo em atuar como um filtro de *trigger online* robusto e altamente discriminativo.

3 Metodologia

Este capítulo detalha a arquitetura computacional e as etapas de processamento desenvolvidas para a deteção de anomalias nas imagens do experimento CYGNO. O algoritmo foi integralmente implementado na linguagem Python, utilizando a biblioteca PyTorch para a construção, treino e inferência dos modelos de aprendizagem profunda (Marchese, 2026). A *pipeline* foi estruturada em quatro etapas principais: pré-processamento das imagens, processamento pelo *Autoencoder* Convolucional, otimização do modelo e extração da Região de Interesse (ROI) pós-processamento.

3.1 Pré-processamento dos Dados

As imagens óticas utilizadas neste trabalho são capturadas por sensores sCMOS individuais que possuem, em sua forma bruta (*RAW*), uma resolução nativa de 2304×2304 pixels e uma profundidade de cor de 16 *bits*, o que permite representar 65535 níveis de intensidade luminosa (Amaro, 2026). Processar imagens com esta magnitude de resolução diretamente numa Rede Neural Convolucional exigiria um custo computacional e alocação de memória de vídeo (VRAM) incompatíveis com as restrições de tempo real de um sistema de *trigger online*. Este problema tende a se agravar em futuras expansões do experimento, como o demonstrador CYGNO-04, cujas exposições combinadas atingirão resoluções de 4096×2304 pixels a altas taxas de aquisição (Amaro, 2026).

Para otimizar o fluxo de dados, a primeira etapa da *pipeline* consiste na aplicação de um recorte de caixa fiducial (*Fiducial Box*). Este corte é posicionado nas coordenadas iniciais $(x, y) = (375, 375)$ e possui dimensões de 1525×1525 pixels, estendendo-se, portanto, até as coordenadas $(1900, 1900)$ da imagem original (Amaro, 2026). O objetivo deste recorte é descartar as bordas extremas do sensor, que frequentemente contêm anomalias estáticas, resposta não-uniforme e ruídos de alta amplitude que não possuem valor para a física do experimento.

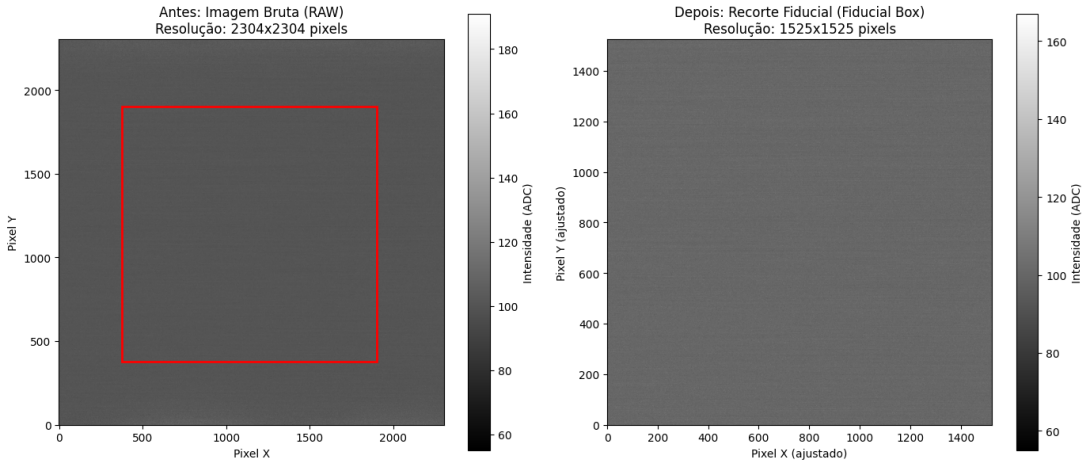


Figura 6: Diagrama esquemático do recorte de caixa fiducial implementado.

Fonte: O autor.

Após o corte, a imagem é submetida a um redimensionamento (*resize*) para as dimensões alvo de 1024×1024 pixels. Este processo é realizado por meio de interpolação bilinear (utilizando o método `INTER_LINEAR`), o qual calcula os novos valores dos pixels a partir da média ponderada dos quatro pixels vizinhos mais próximos. Essa transformação compatibiliza a resolução espacial nativa com a geometria de entrada exigida pela Rede Neural Convolucional.

3.2 Arquitetura do *Autoencoder* Convolucional

A arquitetura central do sistema foi concebida utilizando os módulos da biblioteca PyTorch (`torch.nn`). Distanciando-se de um *Autoencoder* sequencial simples, a topologia da rede adota uma estrutura de compressão e expansão espacial inspirada na arquitetura U-Net, incorporando conexões de salto (*skip connections*) para preservar detalhes morfológicos de alta frequência do ruído de fundo. O fluxo de processamento é dividido em três blocos funcionais: *Encoder*, Gargalo (*Bottleneck*) e *Decoder*.

O bloco construtivo base da rede (`ConvBlock`) é formado por uma camada convolucional 2D com *kernel* de 3×3, seguida por uma normalização de lote (`BatchNorm2d`) e uma função de ativação não-linear ReLU.

O ***Encoder*** é responsável por extrair os mapas de características espaciais da imagem de entrada de 1024×1024 pixels. Em vez de utilizar operações de *max-pooling* tradicionais, a redução de dimensionalidade é realizada pelo módulo `DownBlock`, que aplica convoluções com passo igual a dois (*stride* $= 2$). A cada nível de profundidade, a dimensão espacial dos tensores é reduzida à metade (de 1024 para 512, 256 e, finalmente, 128 pixels), enquanto

a profundidade dos canais de características é aumentada iterativamente (1, 22, 44, 66 e 88 canais). Durante o fluxo de propagação (*forward pass*), os tensores resultantes de cada escala espacial são armazenados em memória para posterior concatenação.

O **Gargalo (*Bottleneck*)** atua como a restrição máxima de informação da rede. O mapa de características final do *Encoder* sofre uma última redução espacial para 64×64 pixels e é submetido a uma operação de Agrupamento Médio Global (*Global Average Pooling* - GAP). O vetor latente resultante alimenta duas camadas lineares totalmente conectadas (`nn.Linear`), que comprimem a informação estocástica da imagem num espaço latente altamente denso de apenas 128 dimensões. Em seguida, a segunda camada densa projeta e remodela este vetor para semear o processo de reconstrução num tensor de base com profundidade de 3 canais e resolução de 64×64 pixels.

O ***Decoder*** realiza a expansão gradual da topologia bidimensional. A ampliação espacial é conduzida por camadas de convolução transposta (`nn.ConvTranspose2d`), que duplicam a resolução geométrica a cada etapa. Imediatamente após cada ampliação, o tensor expandido é concatenado em profundidade com o mapa de características correspondente armazenado previamente pelo *Encoder* (através das *skip connections*). Esta concatenação injeta as texturas espaciais de alta resolução diretamente no processo de decodificação, permitindo que a rede reconstrua o pedestal sem perda de nitidez geométrica.

Por fim, após recuperar a escala original de 1024×1024 pixels, a última camada aplica uma convolução seguida por uma função de ativação *Sigmoid*. Esta etapa assegura o colapso profundo dos canais para um único canal em tons de cinza e restringe as intensidades luminosas reconstruídas ao intervalo normalizado contínuo $[0, 1]$.

3.3 Treinamento e Pós-processamento

O treinamento do modelo foi estruturado sob o paradigma de aprendizado auto-supervisionado (*Self-Supervised Learning* - SSL), utilizando uma arquitetura adaptada de *Denoising Autoencoder* (DAE). Esta abordagem fundamenta-se nos princípios de extração de características robustas e supressão de ruído introduzidos por Vincent et al. (Vincent *et al.*, 2008). Neste contexto, a otimização dos pesos é guiada pelo algoritmo Adam, minimizando uma função de perda que pondera o Erro Quadrático Médio (MSE) e o Índice de Similaridade Estrutural (SSIM), garantindo a estabilidade energética e a preservação morfológica das texturas do ruído original.

A diferença metodológica desta etapa, que caracteriza o sistema como auto-supervisionado, reside na injeção de anomalias sintéticas durante o processo de treinamento. Na literatura de visão computacional, a corrupção artificial de amostras normais para forçar a rede a aprender a restaurá-las ou segmentá-las tem se provado uma técnica eficaz para a detecção de anomalias, dispensando a necessidade de anotações manuais rigorosas (Turowski *et al.*, 2022; Li *et al.*, 2021; Zavrtanik; Kristan; Skočaj, 2021). Para garantir maior eficiência no experimento CYGNO, o modelo precisa ser incapaz de reconstruir trilhas de ionização (possíveis eventos de sinais). Para forçar esse comportamento, o algoritmo gera traços artificiais que são, então, sobrepostas às imagens puras de ruído.

3.3.1 Modelagem Matemática das Anomalias Sintéticas

Para que a rede não aprenda a reproduzir assinaturas físicas do detetor, a morfologia das anomalias injetadas foi projetada para emular duas topologias de eventos: rastros estendidos (gerados por partículas ionizantes) e deposições pontuais ou aglomerados (*blobs*),

frequentemente associados a ruídos intrínsecos de alta amplitude do sensor sCMOS ou a recuos de curtíssimo alcance (Ereditato, 2022; Amaro, 2026).

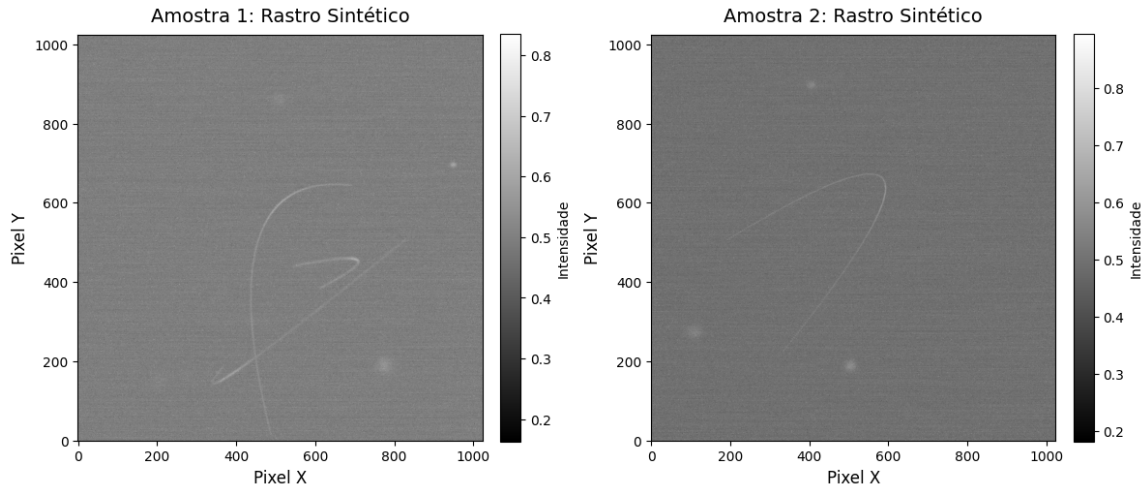


Figura 7: Amostras visuais da injeção de anomalias sintéticas geradas algoritmicamente e sobrepostas ao ruído de pedestal do detetor. Evidenciam-se as duas topologias emuladas: os rastros estendidos com modulação de amplitude e as deposições pontuais.

Fonte: O autor.

O bloco construtivo fundamental de qualquer anomalia no algoritmo é a deposição espacial de energia modelada por uma função Gaussiana bidimensional isotrópica. Dado um centroide (x_c, y_c), a intensidade luminosa depositada $I(x, y)$ num pixel adjacente é definida por:

$$I_{gauss}(x, y) = A \cdot \exp\left(-\frac{(x - x_c)^2 + (y - y_c)^2}{2\sigma^2}\right) \tag{7}$$

onde A representa a amplitude máxima de intensidade luminosa local e σ o desvio padrão transversal, que simula a difusão espacial dos elétrons no gás da TPC. Para otimização computacional, o algoritmo calcula esta deposição apenas numa vizinhança restrita de raio $r = 3\sigma$, região sobre a qual é definida a máscara binária de penalização para o cálculo da função de perda.

Para a geração dos rastros contínuos, a trajetória da partícula é aproximada geometricamente por uma Curva de Bézier Quadrática. Sendo $\mathbf{P}_0$, $\mathbf{P}_1$ e $\mathbf{P}_2$ vetores bidimensionais independentes amostrados aleatoriamente no espaço da imagem, a posição de deposição $\mathbf{P}(u)$ obedece à equação:

$$\mathbf{P}(u) = (1 - u)^2\mathbf{P}_0 + 2(1 - u)u\mathbf{P}_1 + u^2\mathbf{P}_2, \quad \text{para } u \in [0, 1] \tag{8}$$

Para reproduzir o fenômeno físico da perda de energia estocástica ao longo da trajetória de uma partícula (como a flutuação do pico de Bragg), a amplitude do rastro não é constante. A variável de amplitude ao longo da curva, $A(u)$, é dinamicamente modulada por um perfil Gaussiano unidimensional:

$$A(u) = A_0 \left[1 + s \cdot \exp\left(-\frac{1}{2}\left(\frac{u - u_0}{0.08}\right)^2\right)\right] \tag{9}$$

Nesta equação, A_0 é a amplitude base, s dita a intensidade da flutuação (*scaling*) e u_0 define a posição normalizada do pico de energia na trajetória. Durante a execução, a curva é discretizada em 300 passos iterativos, aplicando-se a deposição I_{gauss} centrada em $\mathbf{P}(u)$ com amplitude $A(u)$ a cada incremento (Marchese, 2026).

Por fim, no processo de montagem (*batching*), cada imagem recebe a sobreposição aditiva linear destas componentes sintéticas, aleatoriamente entre 1 a 4 rastros e 3 a 5 aglomerados pontuais, com as intensidades finais truncadas ao valor de 1.0.

3.3.2 Fluxo de Propagação e Penalização Espacial

Durante o fluxo de propagação (*forward pass*), a rede recebe como entrada o tensor corrompido (ruído original + rastro sintético), conforme ilustrado na Figura 8. Contudo, o alvo (*target*) da reconstrução exigido pela função de perda é exatamente a imagem de ruído original limpa. Ao utilizar a máscara espacial associada à anomalia para penalizar severamente a reconstrução nessa região, o modelo é forçado a suprimir qualquer estrutura linear que se assemelhe a um evento físico.

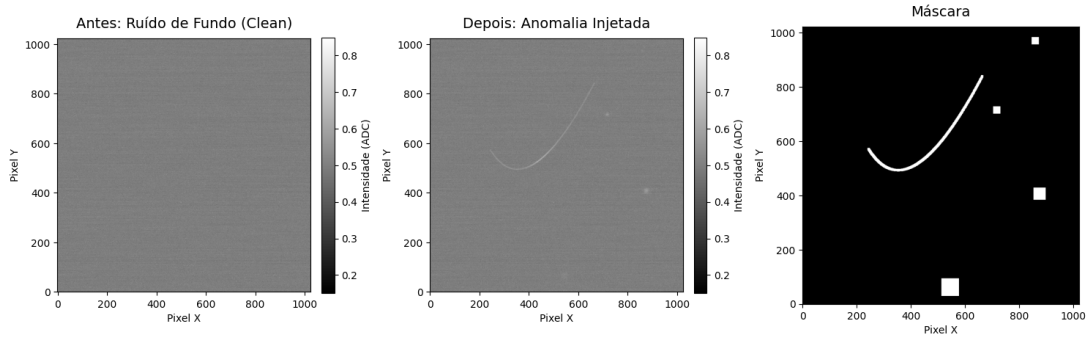


Figura 8: Comparação visual do processo de injeção de sinais sintéticos. À esquerda, apresenta-se uma amostra contendo apenas o ruído de fundo. No centro, a mesma imagem após a injeção de anomalias simuladas, utilizada para compor o conjunto de dados corrompidos para o treino do *Autoencoder*. À direita, a máscara utilizada para penalização durante o treinamento.

Fonte: O autor.

Dessa forma, a rede é treinada para atuar como um filtro cego aos sinais, mapeando o espaço latente exclusivamente para a distribuição estatística do ruído. A etapa de validação é executada inteiramente sobre tensores sem injeções, assegurando que a capacidade de reproduzir o fundo estacionário permaneça intacta ao longo das épocas.

3.3.3 Inferência e Extração da Região de Interesse (ROI)

A estratégia de treinamento descrita otimiza diretamente a fase de inferência e a extração final da Região de Interesse (ROI). Durante a operação de *triggering*, quando uma matriz real x contendo a passagem de uma partícula é processada pela rede, o *Autoencoder* suprime ativamente o rastro bidimensional e devolve uma reconstrução x' composta majoritariamente pelo ruído de fundo estimado.

Ao calcular o mapa de resíduos absolutos $R = |x - x'|$, a ausência proposital do rastro na reconstrução gera um pico de erro de alta magnitude localmente elevado exatamente

nas coordenadas da trajetória física da partícula. Para isolar o evento e transformá-lo numa representação booleana, aplica-se o limiar de binarização empírico estabelecido pelo modelo de referência da colaboração, fixado em $\tau = 0,04$ (Amaro, 2026). Os pixels cujo erro absoluto de reconstrução superam este patamar são preservados como candidatos a sinal anômalo, enquanto o ruído subjacente restante é zerado.

Contudo, devido a flutuações estatísticas locais da cintilação no gás He/CF_4 e à própria resposta estocástica do sensor sCMOS, o mapa binarizado resultante pode apresentar fragmentações e lacunas internas ao longo do rastro. Para solucionar essa descontinuidade morfológica sem comprometer a integridade geométrica do sinal, o algoritmo conduz uma operação de fechamento morfológico (*Morphological Closing*) (Amaro, 2026).

Seguindo os critérios de reconstrução rápida estabelecidos pela colaboração (Amaro, 2026), esta operação utiliza um elemento estruturante elíptico com um raio de vizinhança parametrizado em 40 pixels. Esta janela espacial de fechamento delimita uma matriz estruturante de 81×81 pixels atuando ao redor de cada pixel anômalo detectado. O filtro atua interconectando agrupamentos de carga adjacentes e preenchendo vazios intra-rastro originados pela difusão longitudinal ou transversal dos elétrons de deriva, consolidando uma máscara binária contínua e coesa que delimita a extensão macroscópica real do evento.

O resultado final desta operação gera uma máscara compacta que viabiliza a aplicação do princípio de *zero-suppression* espacial, conforme a figura 9. Toda a vasta região de fundo inativo da matriz original é sumariamente descartada, reduzindo drasticamente a dimensionalidade dos dados e garantindo que apenas o conteúdo contido dentro da ROI estruturada seja serializado e mantido para armazenamento e análise *offline*.

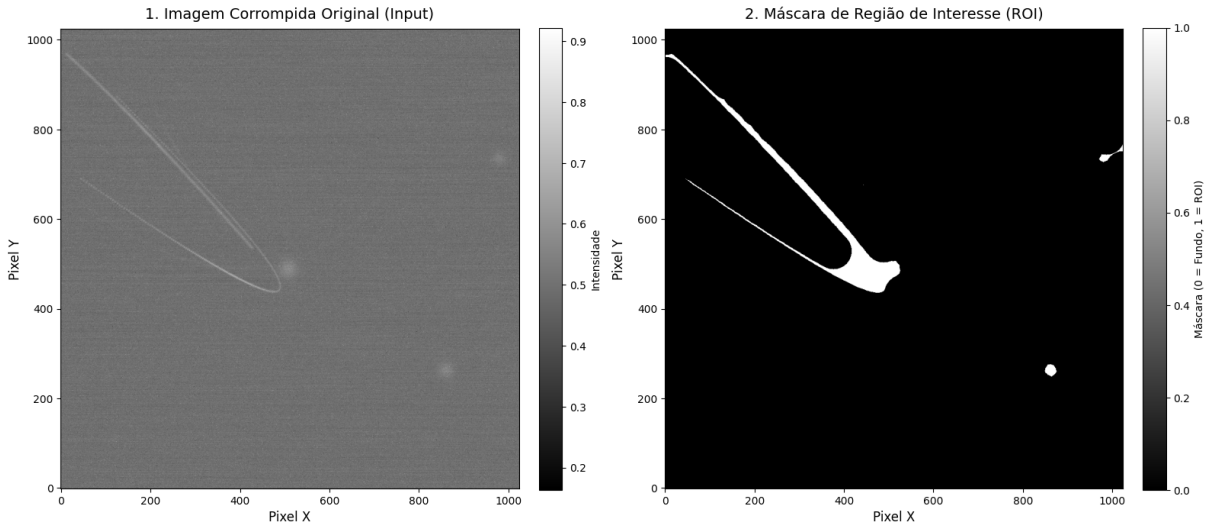


Figura 9: Demonstração visual do impacto do processamento morfológico na compressão de dados (*zero-suppression*). À esquerda, a imagem integral de entrada corrompida pelo ruído de pedestal e contendo anomalias sintéticas (rastros e aglomerados pontuais). À direita, a máscara binária final de Região de Interesse (ROI) isolada pelo algoritmo. Toda a extensão escura da matriz (valor nulo) é ativamente descartada do armazenamento, retendo exclusivamente a área útil do sinal físico.

Fonte: O autor.

4 Resultados e Discussões

Este capítulo apresenta e analisa os resultados obtidos a partir da aplicação do *Auto-encoder* Convolucional como mecanismo de *trigger online* para o experimento CYGNO. Inicialmente, detalha-se a composição do conjunto de dados (*dataset*) utilizado para o treinamento e validação do algoritmo. Em seguida, avalia-se a capacidade do modelo em reconstruir o ruído estacionário do detetor e a sua eficácia no isolamento de trajetórias anômalas, culminando na análise quantitativa do ganho de Relação Sinal-Ruído (SNR) proporcionado pelo filtro.

4.1 Descrição do Conjunto de Dados

Para assegurar a validade física e a aplicabilidade da rede neural aos requisitos do experimento CYGNO, o conjunto de dados empregado neste trabalho possui uma composição híbrida. Para a modelagem do fundo estacionário durante a fase de treinamento, o modelo foi alimentado estritamente com imagens reais de ruído adquiridas experimentalmente pelo protótipo LIME (*Long Imaging ModulE*). Por outro lado, os dados utilizados na fase de teste, que contêm as assinaturas visuais dos eventos físicos de interesse, baseiam-se em imagens geradas por simulação Monte Carlo utilizando o *software* GEANT4, parametrizado com as propriedades do mesmo protótipo (Pains, 2024).

De acordo com os estudos desenvolvidos na colaboração, o experimento tem como foco a detecção de eventos raros e de energia extremamente baixa (Amaro, 2026; Pains, 2024). Sendo assim, os dados utilizados para a avaliação do algoritmo são provenientes de simulações de partículas incidentes com energia restrita a 1 keV (Pains, 2024). Existem duas topologias de eventos físicos de interesse do projeto: recuos nucleares (NR), tipicamente associados a potenciais interações de matéria escura (como WIMPs), e recuos eletrônicos (ER), que representam o ruído de fundo provocado por radioatividade inerente (Pains, 2024). O *dataset* utilizado neste trabalho faz uso somente dos eventos de recuos eletrônicos (ER).

Em sua forma bruta, cada quadro de simulação é gerado como uma matriz de 2304×2304 pixels (Pains, 2024). O processo computacional emula a detecção pelo fotossensor, sobrepondo o sinal isolado da partícula a um fundo de puro ruído eletrônico adquirido de forma empírica pelo detetor (Pains, 2024). Visando submeter o algoritmo de *trigger* às condições mais severas de baixa relação sinal-ruído, o conjunto original de 1 keV serviu também como base para a derivação de simulações com energias de frações de keV, nomeadamente 0.25 keV, 0.5 keV e 0.75 keV (Pains, 2024).

Conforme estabelecido na *pipeline* de pré-processamento detalhado na Metodologia, todas as imagens deste *dataset* são submetidas ao recorte da caixa fiducial, resultando em uma região útil de 1525×1525 pixels, seguida por um redimensionamento bilinear para tensores de 1024×1024 pixels.

Para suportar o paradigma de aprendizado não-supervisionado e permitir uma avaliação robusta, os dados foram particionados da seguinte maneira:

- **Conjunto de Treinamento:** Composto pelas matrizes de ruído de fundo original do sensor (*pedestal*). Este subconjunto utiliza o mecanismo de injeção sintética de anomalias atrelado ao cálculo da função de perda com a imagem limpa como alvo, forçando o modelo a aprender apenas as variáveis latentes do ruído estacionário.

- **Conjunto de Validação:** Composto por quadros isolados estritamente limpos (sem injeções sintéticas nem simulações), inferidos ao final de cada época para atestar

a estabilidade e a convergência das funções de Erro Quadrático Médio (MSE) e similaridade estrutural (SSIM).

- **Conjunto de Teste:** Composto ativamente pelas simulações GEANT4 de 1 keV (ER) e suas derivações de sub-keV (0.25, 0.5 e 0.75 keV) (Pains, 2024). Este conjunto de teste é submetido à inferência da rede exclusivamente ao fim do ciclo de aprendizado para a extração do mapa de resíduos, aplicação do limiar de isolamento ($\tau = 0.04$) e quantificação matemática da eficácia da rede por meio do Ganho de SNR.

4.2 Avaliação Morfológica do Sinal Extraído via SSIM Local

A métrica definitiva para comprovar a eficácia de um filtro de *trigger online* não reside apenas em detetar a presença de uma anomalia, mas na capacidade de extraí-la preservando a sua exata geometria física para análises posteriores de reconstrução de energia e direcionalidade. Para validar quantitativamente esta preservação morfológica, monitorizou-se o Índice de Similaridade Estrutural (SSIM) calculado exclusivamente sobre a Região de Interesse (ROI) do evento.

Durante o processo de inferência do conjunto de teste, a rede recebe a imagem corrompida pelo ruído (x) e gera a sua reconstrução (x'). O sinal extraído pelo algoritmo é definido pelo mapa de resíduos $R = x - x'$. Para aferir a fidelidade desta extração, o resíduo R é comparado diretamente contra o sinal puro original simulado no GEANT4 (S_{true}), o qual atua como o gabarito absoluto (*ground truth*).

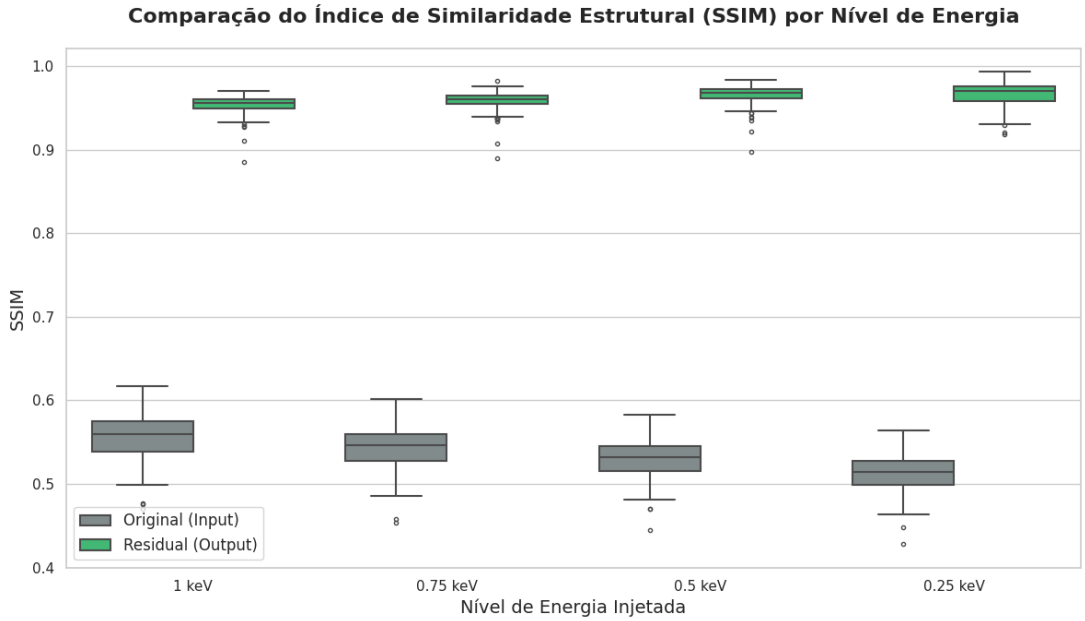


Figura 10: Evolução do Índice de Similaridade Estrutural (SSIM) local, medido exclusivamente na Região de Interesse, entre o rastro extraído pelo mapa de resíduos e o sinal físico original (GEANT4), segregado por faixas de energia.

Fonte: O autor.

Para garantir o rigor estatístico da métrica e impedir que a área de fundo estático

dominasse o cálculo, a avaliação do SSIM foi restrita ao redor do sinal. O algoritmo realiza um corte em volta das coordenadas que envolvem o rastro com um preenchimento de segurança (*padding*) de 5 pixels em todas as direções. A Figura 10 ilustra a evolução deste SSIM restrito ao longo do treinamento para diferentes níveis de energia.

A análise da Figura 10 comprova o sucesso da arquitetura DAE em atuar como um filtro subtrativo não-destrutivo. O crescimento assintótico e a estabilização das curvas indicam que o *Autoencoder* aprende a modelar e subtrair o ruído de fundo sem atenuar severamente os pixels que compõem o rastro da partícula.

Como esperado pelas leis de flutuação estatística de fótons, os eventos de maior energia (1.00 keV e 0.75 keV) atingem os maiores níveis de similaridade, visto que possuem uma densidade de ionização mais robusta e menos suscetível à fragmentação pela subtração do pedestal. Contudo, é notável a resiliência do modelo frente aos recuos tênues de 0.25 keV; mesmo no limite da sensibilidade do detetor sCMOS, o mapa de resíduos consegue reter uma correlação estrutural significativa com a verdade de base (*ground truth*), comprovando que a topologia física da interação original é fortemente preservada pela rede.

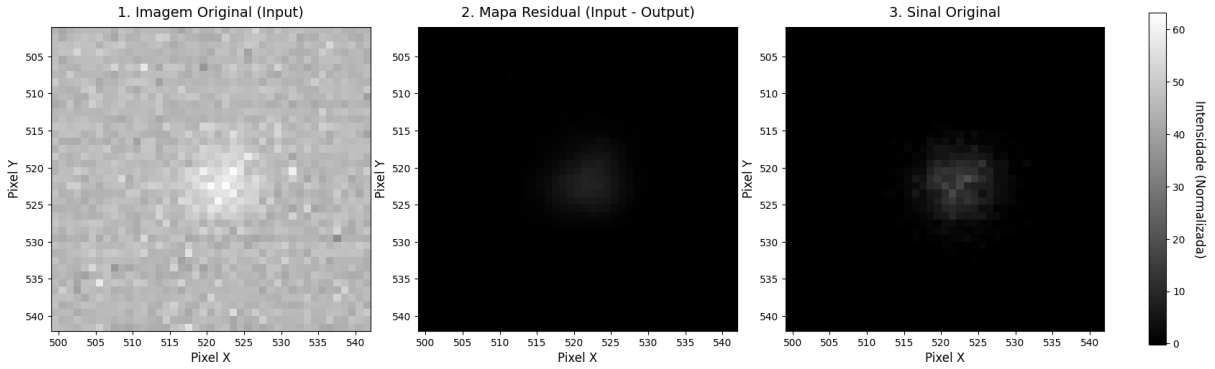


Figura 11: Comparativo visual detalhado na Região de Interesse (ROI) para um evento de 1keV simulado. À esquerda, a imagem de entrada (sinal + ruído). Ao centro, o rastro extraído pelo mapa de resíduos após a inferência do *Autoencoder*. À direita, o sinal físico original.

Fonte: O autor.

4.3 Análise do Ganho de Relação Sinal-Ruído (SNR)

No contexto da instrumentação em física de altas energias, uma métrica muito importante é a Relação Sinal-Ruído (SNR). O detetor CYGNO, ao operar na busca por recuos nucleares de baixa energia, lida intrinsecamente com sinais cuja amplitude frequentemente se confunde com as flutuações estatísticas do pedestal da câmara sCMOS. Dessa forma, o algoritmo proposto deve provar que a operação subtrativa do *Autoencoder* suprime a variância do ruído de forma muito mais agressiva do que atenua a intensidade do sinal da partícula.

Para quantificar esta melhoria de forma absoluta, utiliza-se a métrica do Ganho de SNR ($\mathrm{SNR}_{\mathrm{gain}}$). Conforme estabelecido na fundamentação teórica, a SNR de uma imagem é classicamente definida pela razão entre a intensidade média dos pixels pertencentes à

região do sinal (μ_{sinal}) e o desvio padrão dos pixels na região de fundo estacionário (σ_{fundo}). O ganho de processamento é, portanto, a razão entre a SNR calculada no mapa de resíduos final ($R = |x - x'|$) e a SNR medida na imagem bruta original (x):

$$\text{SNR}_{\text{gain}} = \frac{\text{SNR}_{\text{final}}}{\text{SNR}_{\text{inicial}}} = \frac{(\mu_{\text{sinal}}/\sigma_{\text{fundo}})_{\text{resíduo}}}{(\mu_{\text{sinal}}/\sigma_{\text{fundo}})_{\text{original}}} \tag{10}$$

Valores de $\text{SNR}_{\text{gain}} > 1$ certificam que o processamento resultou em uma imagem com maior destaque para o sinal. A Figura 12 apresenta a distribuição do ganho de SNR obtido durante a fase de inferência no conjunto de teste, segregada pelas energias simuladas de 0.25, 0.50, 0.75 e 1.00 keV.

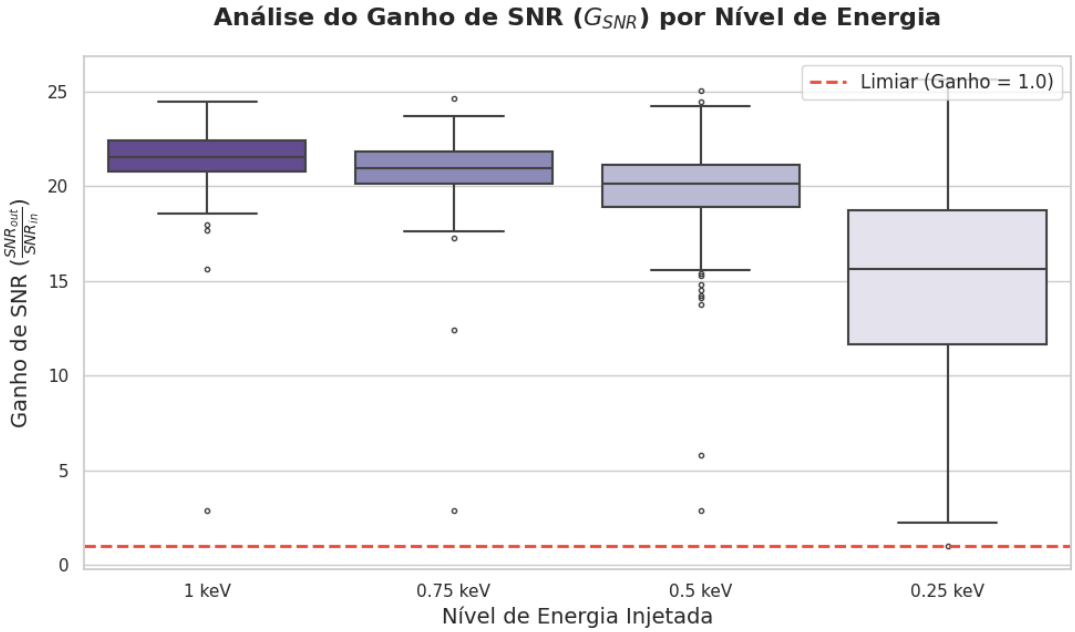


Figura 12: Distribuição do Ganho de Relação Sinal-Ruído (SNR_{gain}) alcançado pelo algoritmo para diferentes faixas de energia de recuo simuladas no conjunto de teste.

Fonte: O autor.

A avaliação dos dados revela o impacto direto da arquitetura na capacidade de deteção do experimento. Observa-se que, para todas as faixas de energia analisadas, o SNR_{gain} posiciona-se consistentemente acima da unidade, validando a eficácia global do filtro em amplificar a detectabilidade dos eventos.

Nas simulações de 1.00 keV, onde o sinal original já possui uma distinção razoável em relação ao ruído de fundo, esperava-se que a rede fosse capaz de aumentar ainda mais a relação sinal ruído. Contudo, em eventos críticos de 0.25 keV, onde o rastro da partícula está quase inteiramente submerso no ruído de leitura do sensor (apresentando uma $\text{SNR}_{\text{inicial}}$ limítrofe), também notou-se um bom desempenho do *Autoencoder*. O ganho de SNR positivo nessa baixa faixa energética comprova a robustez da técnica aplicada.

4.4 Análise da Região de Interesse (ROI) e Compressão de Dados

A validação definitiva da viabilidade do algoritmo como um filtro de *trigger online* reside na sua capacidade de maximizar a rejeição do ruído de fundo sem comprometer a topologia do sinal físico. Para quantificar essa eficiência, conduziu-se uma análise em larga escala focada na métrica de Fração de Área Retida. Esta métrica mensura o percentual de pixels ativos na máscara binária final em relação à área total da imagem original (1024×1024 pixels).

A fim de avaliar a robustez do modelo contra alarmes falsos, o conjunto de teste foi rigorosamente segregado. O algoritmo processou iterativamente amostras contendo exclusivamente ruído e amostras contendo a injeção de anomalias sintéticas. O procedimento foi repetido para os quatro níveis de energia (1.00, 0.75, 0.50 e 0.25 keV). Em todas as inferências, os parâmetros morfológicos foram mantidos constantes, utilizando o limiar de binarização $\tau = 0,04$ e um elemento estruturante elíptico com raio de 40 pixels para o agrupamento espacial. A Figura 13 ilustra a distribuição estatística das frações de área retidas.

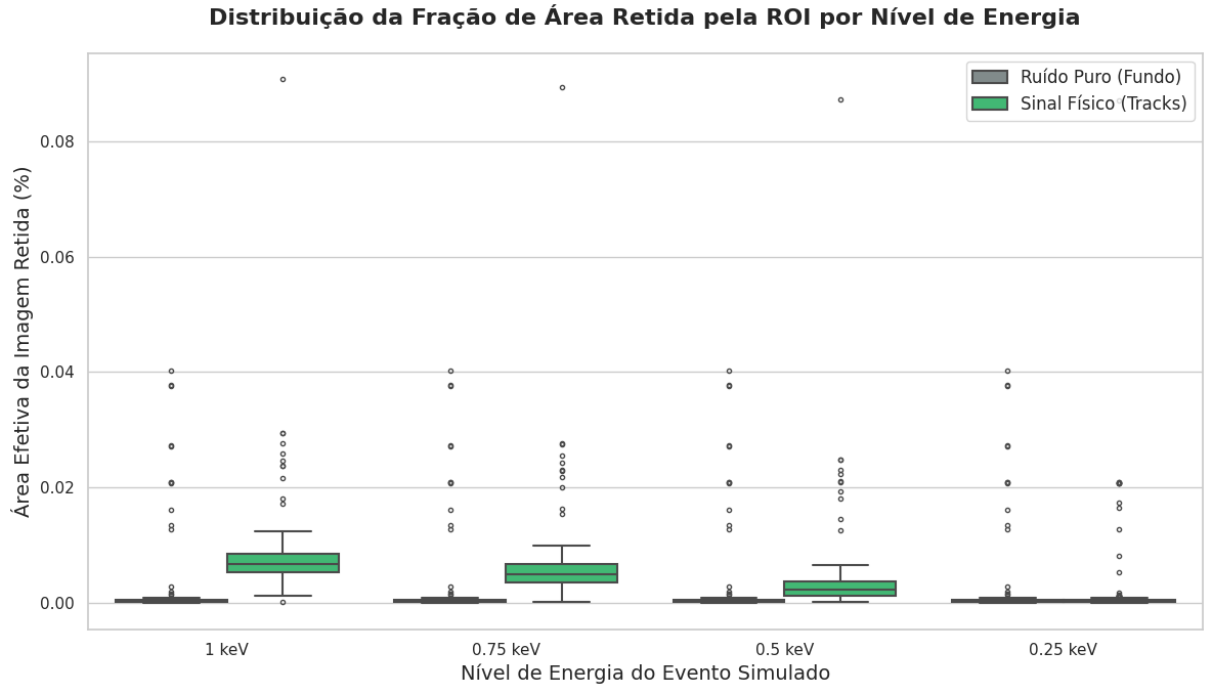


Figura 13: Diagrama de caixas (*boxplot*) comparando a Fração de Área Retida pela máscara de ROI. A análise contrapõe o processamento de matrizes restritas ao ruído de leitura (em cinza) com quadros contendo sinais físicos válidos (em verde), estratificados por energia. Observa-se que a rede rejeita ativamente o pedestal, convergindo o ruído puro para uma área retida nula. Para quadros com sinal, a área ativa isolada pelo algoritmo é tipicamente inferior a 5%, garantindo compressão extrema.

Fonte: O autor.

A interpretação da Figura 13 comprova o sucesso da operação de *zero-suppression*. No que tange aos quadros compostos puramente pelo fundo estacionário, a operação subtrativa do *Autoencoder* neutraliza as flutuações do sCMOS, impedindo que os resíduos

ultrapassem o limiar τ. Como resultado, a área retida para essas amostras converge rigorosamente para zero. Do ponto de vista de aquisição de dados (DAQ), esse comportamento atesta uma formidável taxa de rejeição de fundo, eliminando alarmes falsos e impedindo o armazenamento de "lixo digital".

Por outro lado, as distribuições referentes aos quadros com eventos físicos reais demonstram que a rede recorta e isola a anomalia com sucesso. O comportamento da máscara reflete a natureza intrínseca da deposição de energia: anomalias de 1.00 keV, que geram rastros mais longos e difusos, exigem uma ROI marginalmente maior. À medida que a energia decai, a extensão do evento diminui e a máscara se adapta, retendo frações espaciais ainda menores.

5 Conclusão e Trabalhos Futuros

Este trabalho teve como objetivo principal validar a aplicação de uma arquitetura de *Denoising Autoencoder* (DAE) sob o paradigma de aprendizado auto-supervisionado como um mecanismo de *trigger* online para o experimento CYGNO. O desafio central consistia em mitigar o gargalo de aquisição de dados (DAQ) gerado pelos sensores sCMOS, desenvolvendo um filtro computacional capaz de rejeitar o fundo ruidoso estacionário em tempo real, sem sacrificar a integridade de eventos físicos raros e de baixa energia.

Ao contrário de implementações anteriores, a topologia foi retreinada integralmente a partir do zero, o que permitiu consolidar uma *pipeline* própria de processamento de tensores e comprovar a estabilidade de convergência do modelo frente à distribuição estocástica do ruído. Através da injeção de anomalias sintéticas durante o treinamento e da otimização conjunta pelas métricas de Erro Quadrático Médio (MSE) e Índice de Similaridade Estrutural (SSIM), o *Autoencoder* aprendeu a omitir ativamente assinaturas morfológicas de sinal, atuando como um eficiente filtro subtrativo não-destrutivo na fase de inferência.

Os resultados quantitativos obtidos no conjunto de teste com simulações GEANT4 atestam a robustez do algoritmo nos seguintes aspectos:

- **Preservação Morfológica:** A análise do SSIM local demonstrou que a geometria dos rastros foi fielmente preservada no mapa de resíduos absolutos. Mesmo para eventos limítrofes na faixa de 0.25 keV, que operam no limite da sensibilidade do detetor, a rede conseguiu reter uma correlação estrutural significativa com a verdade de base (*ground truth*).

- **Amplificação do Sinal:** O Ganho de Relação Sinal-Ruído (SNR_{gain}) manteve-se consistentemente superior a 1 em todas as faixas de energia analisadas (0.25 keV a 1.00 keV). Este indicador quantifica que o processamento suprimiu a variância do ruído de leitura de forma muito mais agressiva do que atenuou a intensidade luminosa das trajetórias físicas.

- **Compressão de Dados (Zero-Suppression):** A etapa de pós-processamento baseada em limiarização e fechamento morfológico validou a viabilidade do filtro para a compressão espacial. A avaliação da Fração de Área Retida evidenciou que amostras de ruído puro convergiram rigorosamente para uma área nula, suprimindo o armazenamento de alarmes falsos. Para quadros contendo sinais válidos, a máscara reteve tipicamente menos de 5% da extensão original da matriz, isolando estritamente a Região de Interesse (ROI).

Contudo, antes de concluir definitivamente que todos os critérios operacionais foram plenamente atendidos para a integração na linha de produção do experimento, ressalta-se a necessidade imperativa de realizar uma análise sistemática do tempo de execução do algoritmo. Dado que a latência de processamento por quadro é um fator essencial e limitante para sistemas de *trigger* online, a quantificação exata do tempo de inferência e do *throughput* de tensores constitui uma etapa mandatória para assegurar o cumprimento dos requisitos de tempo real do detetor.

Como desdobramentos e continuidade deste estudo, propõem-se os seguintes trabalhos futuros:

- Realizar um mapeamento temporal detalhado de cada etapa da *pipeline* de processamento para validar a viabilidade de operação na taxa de aquisição nominal do experimento;
- Avaliar o comportamento e a eficiência de detecção do algoritmo utilizando dados de sinais reais coletados diretamente pelo protótipo, confrontando as métricas obtidas nas simulações com a resposta física real e com as dinâmicas estocásticas do detetor;
- Investigar técnicas de otimização computacional, tais como a quantização de pesos e a poda (*pruning*) da rede neural, visando a redução do tempo de execução em hardware dedicado sem degradação do ganho de SNR em regimes de sub-keV.

Dessa forma, embora a validação morfológica e estatística via dados simulados apresente um panorama altamente promissor, a consolidação definitiva desta ferramenta como uma solução escalável para as futuras etapas operacionais do experimento CYGNO, como o demonstrador CYGNO-04, permanece vinculada à execução bem-sucedida destas próximas etapas analíticas e experimentais.

Referências

AMARO, A. Fast reconstruction-based ROI triggering via anomaly detection in the CYGNO optical TPC. **Machine Learning: Science and Technology**, v. 1, p. 1–10, 2026.

AMARO, F. D. *et al.* 3D track reconstruction for the CYGNO experiment. **The European Physical Journal C**, Springer, v. 83, n. 10, p. 937, 2023. DOI: 10.1140/epjc/s10052-023-11988-9. Disponível em: https://link.springer.com/article/10.1140/epjc/s10052-023-11988-9.

BANK, Dor; KOENIGSTEIN, Noam; GIRYES, Raja. Autoencoders. **arXiv preprint arXiv:2003.05991**, 2020.

BARACCHINI, Elisabetta *et al.* **The CYGNO experiment: a directional Dark Matter detector with optical readout**. [*S. l.: s. n.*], 2023. Acesso em: 15 jun. 2026. Disponível em: https://www.researchgate.net/publication/373960384_The_CYGNO_experiment_a_directional_Dark_Matter_detector_with_optical_readout.

CHOLLET, François. **Building Autoencoders in Keras**. The Keras Blog. 2016. Disponível em: https://blog.keras.io/building-autoencoders-in-keras.html. Acesso em: 25 maio 2026.

EREDITATO, Antonio. The CYGNO Experiment. **Instruments**, v. 1, p. 1–10, 2022.

LCTPC COLLABORATION. **Time Projection Chamber**. 2026. Disponível em: https://www.lctpc.org/e8/e57671. Acesso em: 19 maio 2026.

LI, Chun-Liang *et al.* Cutpaste: Self-supervised learning for anomaly detection and localization. *In:* PROCEEDINGS of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [*S. l.: s. n.*], 2021. p. 9664–9674.

MARCHESE, Antonio. **fast_reco_autoencoder: algoritmo de reconstrução rápida e deteção de anomalias por meio de autoencoders para o experimento CYGNO**. Juiz de Fora: [*s. n.*], 2026. Disponível em: https://github.com/antonioMarchese/fast_reconstruction. Acesso em: 27 maio 2026.

PAINS, Igor Fonseca. Desenvolvimento de Algoritmo de Trigger para Detecção de Sinais no Experimento CYGNO. **Universidade Federal de Juiz de Fora**, v. 1, n. 1, p. 1–10, 2024.

TUROWSKI, Marian *et al.* Modeling and generating synthetic anomalies for energy and power time series. *In:* PROCEEDINGS of the Thirteenth ACM International Conference on Future Energy Systems. [*S. l.: s. n.*], 2022. p. 471–484. DOI: 10.1145/3538637.3539760.

VINCENT, Pascal *et al.* Extracting and composing robust features with denoising autoencoders. *In:* ACM. PROCEEDINGS of the 25th international conference on Machine learning. [*S. l.: s. n.*], 2008. p. 1096–1103. DOI: 10.1145/1390156.1390294.

WANG, Zhou prestige *et al.* Image quality assessment: from error visibility to structural similarity. **IEEE Transactions on Image Processing**, IEEE, v. 13, n. 4, p. 600–612, 2004.

ZAVRTANIK, Vitjan; KRISTAN, Matej; SKOČAJ, Danijel. DRAEM: A discriminatively trained reconstruction embedding for surface anomaly detection. *In:* PROCEEDINGS of the IEEE/CVF International Conference on Computer Vision. [*S. l.: s. n.*], 2021. p. 8330–8339.

ZHANG, Aston *et al.* **Dive into Deep Learning**. [*S. l.*]: Cambridge University Press, 2023. https://pt.d2l.ai/.