



UNIVERSIDADE FEDERAL DO RIO GRANDE FURG
CENTRO DE CIÊNCIAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

Tese de Doutorado

GranSeg: Uma abordagem híbrida para segmentação de objetos sobrepostos em problemas de granulometria visual

Nelson de Faria Traversi

Tese de Doutorado apresentada ao PPPGMC da Universidade Federal do Rio Grande, como requisito parcial para a obtenção do grau de Doutor em Modelagem Computacional.

Orientador: Profª. Drª Sílvia Silva da Costa Botelho

Rio Grande, 2025

Ficha Catalográfica

T781g Traversi, Nelson de Faria.
GranSeg: uma abordagem hÍbrida para segmentação de objetos sobrepostos em problemas de granulometria visual / Nelson de Faria Traversi. – 2025.
119 f.

Tese (doutorado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-Graduação em Modelagem Computacional, Rio Grande/RS, 2025.
Orientadora: Dra. Silvia Silva da Costa Botelho.

1. Segmentação 2. Machine Learning 3. Grãos de fertilizante I. Botelho, Silvia Silva da Costa II. Título.

CDU 004

Catalogação na Fonte: Bibliotecário José Paulo dos Santos CRB 10/2344

Nelson de Faria Traversi

“GranSeg: Uma abordagem híbrida para segmentação de objetos sobrepostos em problemas de granulometria visual”

Tese apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal do Rio Grande - FURG, como requisito parcial para obtenção do Grau de Doutor. Área de concentração: Modelagem Computacional.

Aprovado em: 29/04/2025

BANCA EXAMINADORA

gov.br Documento assinado digitalmente
SILVIA SILVA DA COSTA BOTELHO
Data: 24/07/2025 14:38:32-0300
Verifique em https://validar.iti.gov.br

Profa. Dra. Silvia Silva da Costa Botelho (Orientadora - FURG)

Assinado digitalmente por
RODRIGO DA SILVA GUERRA
CPF 941.388.340-87
29 de jul. de 2025

Prof. Dr. Rodrigo da Silva Guerra (FURG)

gov.br Documento assinado digitalmente
EDUARDO TODT
Data: 01/08/2025 19:51:02-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Eduardo Todt (UFPR)

Emanuel da Silva Diaz Estrada:00079780067
Digitally signed by Emanuel da Silva Diaz Estrada:00079780067
DN: CN=Emanuel da Silva Diaz Estrada:00079780067, OU=FURG - Universidade Federal do Rio Grande, O=ICPEdu, C=BR
Date: 2025.07.25 10:50:02-03'00'

Prof. Dr. Emanuel da Silva Diaz Estrada (FURG)

Rio Grande - RS
2025

RESUMO

TRAVERSI, Nelson de Faria. **GranSeg: Uma abordagem híbrida para segmentação de objetos sobrepostos em problemas de granulometria visual**. 2025. 118 f. Tese (Doutorado) – Programa de Pós-Graduação em Modelagem Computacional. Universidade Federal do Rio Grande FURG, Rio Grande.

A segmentação de objetos homogêneos com alto nível de sobreposição e ausência de fundo permanece um desafio significativo no processamento digital de imagens. Com o intuito de investigar soluções para esse problema, este trabalho apresenta uma revisão abrangente dos métodos clássicos de segmentação e das abordagens baseadas em aprendizado de máquina, avaliando suas efetividades em datasets sintéticos e em datasets anotados manualmente. Inicialmente, devido à dificuldade de se obter um conjunto de dados reais devidamente rotulado, desenvolveu-se um simulador de imagens de superfície para gerar um dataset sintético com anotações detalhadas. Esse conjunto de dados possibilitou testes iniciais dos métodos determinísticos na segmentação de granulados, contemplando desde configurações esparsas até cenários com sobreposição total. Observou-se que métodos clássicos (como limiarização e segmentação por contornos) não se mostraram adequados para imagens sem fundo, sobretudo em função da variação de tamanho, formato e coloração homogênea dos grãos, dificultando a escolha de um threshold ideal. Em seguida, para ampliar a robustez dos experimentos, foram criados dois datasets anotados manualmente: um em ambiente laboratorial controlado e outro em ambiente fabril relevante. O objetivo foi comparar o desempenho dos algoritmos em condições mais próximas da aplicação prática, além de avaliar os desafios impostos por ruídos de iluminação, variação de foco e movimentação dos grãos. Considerando a escassez de dados anotados em ambiente real, propôs-se uma metodologia de combinação entre o dataset virtual, o dataset real controlado e o dataset real de ambiente fabril. Essa estratégia permitiu aproveitar o maior número de imagens disponíveis nos dois primeiros datasets para pré-treinamento do modelo e, posteriormente, realizar um fine tuning direcionado pelas imagens capturadas no ambiente fabril. Além disso, foi analisado o impacto de produzir a segmentação em formato binário ou em formato RGB, visando compreender a efetividade de cada abordagem no processo de convergência do modelo. Na etapa seguinte, realizou-se um estudo comparativo de diferentes arquiteturas de redes neurais convolucionais (CNNs) para a substituição do encoder na U-Net, buscando identificar a melhor configuração para lidar com a segmentação de grãos em cenários desafiadores. Os modelos foram avaliados quanto à precisão, sensibilidade e especificidade na detecção dos contornos e na contagem dos objetos segmentados. Por fim, com base nos resultados obtidos no comparativo de diferentes arquiteturas de redes neurais convolucionais (CNNs) e na análise dos métodos clássicos de segmentação,

propôs-se um novo modelo focado na segmentação de objetos sobrepostos em problemas de granulometria visual. Esse modelo faz uso de uma abordagem híbrida, combinando técnicas de segmentação determinística com a imagem original como entradas, o que aumenta a robustez e a precisão em cenários de alta sobreposição. Para validar a proposta, o modelo foi comparado a metodologias consideradas estado da arte, demonstrando eficácia tanto em relação ao número de grãos detectados quanto à consistência das segmentações. Os resultados promissores apontam para a viabilidade de aplicação dessa abordagem em processos industriais, contribuindo para o aprimoramento do controle de qualidade de produtos granulares.

Palavras-chave: Segmentação, Machine Learning, Grãos de fertilizante.

ABSTRACT

TRAVERSI, Nelson de Faria. **GranSeg: A hybrid approach for overlapping object segmentation in visual granulometry problems**. 2025. 118 f. Thesis (Doctorate degree) – Programa de Pós-Graduação em Modelagem Computacional. Universidade Federal do Rio Grande FURG, Rio Grande.

Segmentation of homogeneous objects with a high level of overlap and no background remains a significant challenge in digital image processing. To investigate solutions to this problem, this work presents a comprehensive review of classical segmentation methods and machine learning-based approaches, evaluating their effectiveness on both synthetic and manually annotated datasets. Initially, due to the difficulty of obtaining a properly labeled real data set, a surface image simulator was developed to generate a synthetic dataset with detailed annotations. This data set enabled initial tests of deterministic methods in granule segmentation, covering everything from sparse configurations to scenarios with total overlap. It was observed that classical methods (such as thresholding and contour segmentation) were not suitable for images without a background, mainly due to the variation in size, shape, and homogeneous coloring of the grains, making it difficult to choose an ideal threshold. Next, to increase the robustness of the experiments, two manually annotated datasets were created: one in a controlled laboratory environment and the other in a relevant factory environment. The objective was to compare the performance of the algorithms in conditions closer to practical application, in addition to evaluating the challenges posed by lighting noise, focus variation, and grain movement. Considering the scarcity of annotated data in a real environment, a methodology was proposed to combine the virtual dataset, the real controlled dataset and the real dataset from the factory environment. This strategy made it possible to take advantage of the largest number of images available in the first two datasets to pre-train the model and, subsequently, carry out fine-tuning directed by the images captured in the factory environment. Furthermore, the impact of producing the segmentation in binary format or in RGB format was analyzed, aiming to understand the effectiveness of each approach in the model convergence process. In the next step, a comparative study was carried out on different convolutional neural network (CNN) architectures to replace the encoder in U-Net, seeking to identify the best configuration to deal with grain segmentation in challenging scenarios. The models were evaluated for accuracy, sensitivity and specificity in detecting contours and counting segmented objects. Finally, based on the results obtained in the comparison of different architectures of convolutional neural networks (CNNs) and the analysis of classical segmentation methods, a new model was proposed focused on the segmentation of overlapping objects in visual granulometry problems. This model makes use of a hybrid approach, combining

deterministic segmentation techniques with the original image as inputs, which increases robustness and accuracy in high overlap scenarios. To validate the proposal, the model was compared to methodologies considered state of the art, demonstrating effectiveness both in relation to the number of grains detected and the consistency of the segmentations. The promising results point to the feasibility of applying this approach in industrial processes, contributing to the improvement of product quality control

Keywords: Segmentation, Machine Learning, Fertilizer grains.

LISTA DE FIGURAS

1 Imagem de uma esteira em fábrica de fertilizantes 13
2 Peneiras. 15
3 Agitador de peneiras. 16
4 CAMSIZER para npk e outros granulares [Microtrac 2025] 18

5 Configuração das camadas da rede VGG, variando de 11 até 19 camadas, conforme proposto em sua publicação original. 39
6 Arquitetura da ResNet com conexões residuais (*skip connections*). . . 40
7 Bloco de aprendizado residual do trabalho original onde a conexão X *identity* representa uma *skip connection* [He et al. 2016]. 41
8 Bloco denso de 5 camadas, como apresentado no trabalho original [Huang et al. 2017], demonstrando que cada camada recebe todos os mapas de características anteriores como entrada. 41
9 Arquitetura da U-Net. 43
10 Segmentação perfeita em objetos distintos SAM 49
11 Segmentação limitada em objetos similares 49
12 Segmentação Mask2Former 50
13 Segformer 50
14 Oneformer 51
15 segdiff 52

16 Comparação de imagens: Simulação, Ambiente Controlado e Fábrica. 54
17 Imagem capturada na fábrica. 55
18 Imagem de grãos pré-processados. 56
19 Fluxograma do simulador. 59
20 Imagem RGB real. 60
21 Máscara da imagem real. 61
22 Figura RGB virtual e Máscara 62
23 Ruídos de blur, salt & pepper e gamma alterado em intensidade 0.5 . 63
24 Recorte de 900x900 com redimensionamento para 224x224 da imagem RGB e sua mascara 65
25 Mascara da imagem de fábrica. 66
26 Comparação de brilho, contraste e histograma de cores entre os conjuntos de dados simulado, ambiente controlado e fábrica. 67

27 Output RGB da Unet e Output Binario da unet com Resnet101 encoder 75

28 Modelo GranSeg 81

29 Amostra original 89
30 Amostra com blur com kernel tamanho 3x3 90
31 Amostra com blur com kernel tamanho 5x5 90
32 Amostra com blur com kernel tamanho 9x9 91
33 Amostra com filtro salt & pepper 5% 92
34 Amostra com filtro salt & pepper 10% 92
35 Exemplo de entrada em rgb 100
36 Exemplo de saida esperada 100
37 Comparação visual dos resultados obtidos por diferentes métodos de segmentação sem a presença de ruído nas imagens originais. A primeira linha mostra a máscara esperada (ground truth), seguida pelas saídas dos métodos: Granseg Watershed (segunda linha), DenseNet (terceira linha), VGG19-Unet (quarta linha) e Segment Anything Model (SAM, quinta linha). É possível observar visualmente que o método Granseg Watershed apresenta segmentações com maior integridade estrutural e continuidade nas bordas dos objetos segmentados, em comparação às segmentações mais fragmentadas geradas pelos demais métodos. 102
38 Entrada rgb com filtro blur 3x3 103
39 Comparação visual dos resultados obtidos pelos diferentes modelos após a aplicação do filtro de desfoque (blur 3×3). A primeira linha apresenta a máscara esperada (ground truth), seguida pelas saídas dos métodos Granseg Watershed (segunda linha), DenseNet (terceira linha), VGG19-Unet (quarta linha) e Segment Anything Model (SAM, quinta linha). Observa-se visualmente o impacto do desfoque nas segmentações, sendo possível avaliar qualitativamente a resistência dos métodos em manter a integridade das bordas e estruturas segmentadas. 104
40 Imagens de entrada com filtro de blur de 9x9 105
41 Resultados visuais obtidos após a aplicação do filtro de desfoque mais intenso (blur 9×9). A primeira linha mostra as máscaras esperadas (ground truth), enquanto a segunda e terceira linhas apresentam as segmentações obtidas pelo método Granseg Watershed e pelo Segment Anything Model (SAM), respectivamente. Nota-se o efeito significativo do desfoque sobre os resultados, permitindo avaliar qualitativamente a capacidade dos métodos em manter a coerência estrutural e a continuidade das bordas segmentadas sob alta degradação visual. . 106
42 Entrada RGB com 5% de chance de salt & pepper 107
43 Resultados visuais dos modelos submetidos ao ruído impulsivo do tipo *salt and pepper* com intensidade de 5%. A primeira linha apresenta as máscaras esperadas (ground truth), enquanto as linhas seguintes exibem as segmentações obtidas pelos métodos Granseg Watershed (segunda linha) e Segment Anything Model (SAM, terceira linha). Observa-se qualitativamente o impacto desse ruído na capacidade dos métodos em preservar a continuidade e integridade das estruturas segmentadas. 108

LISTA DE TABELAS

1 Peneiras utilizadas na indústria de fertilizantes. 14

2 Comparação entre métodos *estado da arte* para segmentação 48

3 Resultados experimentais. 71

4 Resultados U-net, Densenet-U e Vgg16-U 76
5 Resultados Vgg19-U, Resnet50-U e Resnet101-U 76

6 Tabela com a comparação de diferentes redes com entrada padrão. . . 95
7 Tabela com a comparação de diferente redes com entrada com blur 3x3 95
8 Tabela com a comparação de diferente redes com entrada com blur 5x5 96
9 Tabela com a comparação de diferente redes com entrada com blur 9x9 97
10 Tabela com a comparação de diferente redes com entrada com 5% salt and pepper 98
11 Tabela com a comparação de diferente redes com entrada com 10% salt and pepper 98

SUMÁRIO

1 Introdução **12**
1.1 Contextualização do problema 14
1.1.1 Controle de Granulometria por peneiramento 15
1.1.2 Controle de granulometria por imagens 16
1.2 Proposta deste trabalho 18
1.3 Motivação 19
1.4 Hipótese Científica 22
1.5 Objetivos 22
1.5.1 Objetivos Específicos 23
1.6 Publicações 23
1.7 Organização da Tese 24

2 Referencial Teórico **26**
2.1 Principais Conceitos 26
2.1.1 Visão Computacional 27
2.1.2 Redes Neurais Convolucionais 27
2.1.3 Janela Deslizante (Sliding Window) 27
2.1.4 Camada Convolucional 27
2.1.5 Camada de Pooling e Downsampling 28
2.1.6 Operações de Upsampling 28
2.1.7 *Thresholding* 29
2.1.8 Equalização de Histograma Adaptativo 29
2.1.9 Segmentação em Visão Computacional 30
2.1.10 Redes Neurais Convolucionais 30
2.1.11 Funcionamento das Camadas Neurais 31
2.1.12 Funções de Ativação 32
2.1.13 Dados Anotados 32
2.1.14 Transfer Learning 32
2.1.15 Skip Connections 33
2.1.16 *Encoder* em Redes Neurais Convolucionais 33
2.1.17 *Decoder* 33
2.1.18 *Encoder-Decoder* 34
2.1.19 *Autoencoder* 34
2.2 Segmentação 35
2.2.1 Segmentação Baseada em Região 35
2.2.2 Segmentação Baseada na Detecção de Bordas 35

2.2.3 Segmentação Baseada em Limiarização (*Thresholding*) 36
2.2.4 Métodos Híbridos e Redes Neurais 37
2.2.5 Revisão dos métodos clássicos apresentados 37
2.3 Segmentação com redes profundas 38
2.3.1 VGG 38
2.3.2 ResNet 39
2.3.3 DenseNet 40
2.3.4 U-Net 42
2.4 Métricas de Avaliação 43
2.4.1 Medidas de Similaridade 44
2.5 Data Augmentation 46
2.6 Trabalhos relacionados a segmentação densa usando algoritmos clássicos 46
2.6.1 Metodos de segmentação estado da arte 48
2.6.2 Considerações finais em relação aos modelos em estado da arte 51

3 Uma metodologia para obtenção de *datasets* de imagens densas com sobreposição de grãos **54**
3.1 Simulador 56
3.2 *Dataset* Manualmente Anotado 60
3.3 Aplicação dos Métodos para a Obtenção dos *Datasets* para o Estudo . . 61
3.3.1 Dataset Sintético 61
3.3.2 Dataset real em ambiente relevante 64
3.4 Comparação de brilho, contraste e equilíbrio de cores 66
3.5 Análise Comparativa da Eficácia de Conjuntos de Dados Sintéticos, Controlados e Reais no Treinamento de Modelos de Segmentação 68
3.5.1 Resultados da Validação 70
3.5.2 Considerações finais em relação a combinação de datasets 73

4 Avaliação Comparativa de *Encoders* na U-Net em conjunto com Impacto de Entradas RGB e Binárias na Segmentação de Imagens **74**
4.1 Dataset 75
4.2 Arquitetura Experimental 75
4.3 Métricas de Avaliação 76
4.4 Resultados e Discussões 76
4.4.1 U-Net 76
4.4.2 DenseNet-U 77
4.4.3 VGG16-U 77
4.4.4 VGG19-U 78
4.4.5 ResNet50 U-Net 78
4.4.6 ResNet101 U-Net 78
4.4.7 Considerações Finais sobre a Comparação de *Encoders* na U-Net e o Impacto das Entradas RGB e Binárias na Segmentação de Imagens 79

5 GranSeg: uma abordagem híbrida para segmentação de objetos sobrepostos no contexto da granulometria visual de fertilizantes **80**
5.1 Overview da Proposta 80
5.1.1 Principais Módulos do modelo 82
5.2 *Setup* e Treinamento 86
5.2.1 Configurações de Treinamento dos Modelos 86

6 Teste e Validação **88**
6.1 Perturbações Relevantes 88
6.1.1 Aplicação de filtro Gaussian para Simulação da Perda de Qualidade das Bordas 89
6.1.2 Aplicação do Ruído *Salt and Pepper* 91
6.2 Avaliação Quantitativa da rede Granseg 93
6.2.1 Análise dos Resultados 94
6.2.2 Considerações finais quanto a sensibilidade do Método a velocidade de esteira 97
6.2.3 Impacto do Ruído Salt and Pepper nos Modelos de Segmentação 97
6.2.4 Considerações finais quanto a sensibilidade do Método a Poeira 99
6.3 Avaliação da Metodologia de Contagem dos Grãos 99
6.3.1 Resultados Iniciais Sem Aplicação de Ruído 100
6.3.2 Resultados com Aplicação de Desfoque (Blur 3x3) 103
6.3.3 Resultados com Aplicação de Desfoque Intenso (Blur 9x9) 105
6.3.4 Resultados com Aplicação de Ruído Salt and Pepper (5%) 107
6.3.5 Considerações 109

7 Conclusão **110**

Referências **112**

1 INTRODUÇÃO

A agricultura é uma das áreas mais importantes para a economia de muitos países [Fuglie et al. 2019], sendo um setor vital para o sustento de toda a humanidade [Sharma et al. 2020]. Atualmente, existe uma preocupação mundial com os impactos do crescimento populacional, que se estima em torno de 9 bilhões no final deste século [Vollset et al. 2020], na capacidade de produção de alimentos [Maja e Ayano 2021], na demanda por energia [48] e na gestão de água potável [Tehreem et al. 2020]. Além disso, as mudanças climáticas exigem novas estratégias de desenvolvimento sustentável que garantam a segurança alimentar, hídrica e energética, preservando o meio ambiente para evitar um colapso global que ponha em risco a existência da humanidade [Cavicchioli et al. 2019].

Uma contribuição para a redução dos impactos ambientais é a transformação da agricultura tradicional em agricultura de precisão. Sabe-se que a intensificação dos métodos agrícolas convencionais tem um conjunto relevante de impactos nocivos, entre eles: oferta excessiva de nitrogênio reativo, eutrofização de terras e corpos d'água, emissões de gases de efeito estufa, perdas de biodiversidade, redução de fluxos superficiais e esgotamento de águas subterrâneas [9, Smith et al. 2013, Steffen et al. 2015, Foster e Custodio 2019]. Além disso, o uso indiscriminado de fertilizantes a longo prazo também implica em perda de eficiência, tornando o solo pobre e levando à redução da produtividade das culturas [Hu et al. 2020]. Por outro lado, a agricultura de precisão visa otimizar a inserção do insumo na lavoura, preservando os recursos [Mavridou et al. 2019]. No entanto, se os insumos não tiverem qualidade, todo o esforço e tecnologia não serão suficientes para garantir o aumento da produção, bem como sua sustentabilidade.

Desse modo, a adoção de práticas de agricultura de precisão exige não apenas a aplicação controlada e direcionada dos insumos, mas também a garantia de que os próprios fertilizantes atendam às exigências de qualidade necessárias para aproveitar todo o potencial dessas tecnologias. É fundamental que as características físico-químicas do produto estejam adequadas para promover a liberação e a absorção eficiente dos nutrientes, mantendo a sustentabilidade e a rentabilidade da produção.

Os fertilizantes possuem certas características químicas e físicas que definem sua efe-

tividade. Uma das características físicas necessárias para aumentar a eficácia dos fertilizantes está relacionada a sua granulometria. Esse fator está diretamente relacionado com a facilidade do fertilizante se misturar ao solo e é um dos principais parâmetros a serem supervisionados ao longo de seu processo de produção industrial, juntamente com as proporções de mistura dos elementos que o compõem. Caso os grãos estejam fora de uma categoria ideal, eles devem ser reprocessados até atingir o volume determinado, para atingir a solubilidade correta [Wang et al. 2021]. Naturalmente, o reprocessamento implica diretamente no tempo e nos custos de produção. Portanto, novas soluções tecnológicas têm sido propostas para superar esse gargalo de produção.

Na indústria, com toneladas de fertilizante são produzidas a cada hora, sendo transportados por esteiras cerca de um metro de largura em formato de V, com profundidade de grãos sobrepostos de 20 a 30 centímetros, semelhantes à esteira da Figura 1. Métodos



Figura 1: Imagem de uma esteira em fábrica de fertilizantes

convencionais de avaliação da uniformidade de grãos de fertilizantes utilizam peneiras mecânicas, onde são analisados o tamanho das suas partículas, se constituindo um processo lento e com necessidade de intervenção humana para ser realizado. Como alternativa processos baseados em sensores utilizando desvio na linha de produção e maquinário especializado, como por exemplo o analisador de partículas CAMSIZER (ver Figura 4). Entretanto o uso desse maquinário é custoso e necessita de adaptações na linha de produção.

Como alternativa aos métodos clássicos de peneiramento mecânico e ao uso de máquinas especializadas com desvio do fluxo de produção, técnicas de visão computa-

cional podem ser aplicadas. Estas abordagens tem como objetivo estimar a granularidade dos grãos a partir do uso de técnicas de visão computacional para segmentar cada grão e estimar suas dimensões.

No entanto o uso de imageamento através de câmeras instaladas na linha de produção traz alguns desafios devido à sobreposição dos grãos implicando na ausência de fundo e de bordas na imagem. A poeira, o movimento e a iluminação também devem ser tratados. Neste trabalho, explora-se a viabilidade do uso de câmeras posicionadas para capturar uma visão superior da linha de produção, bem como a aplicação de métodos de segmentação clássicos e de redes neurais artificiais para realizar a segmentação de forma eficiente.

Criar uma alternativa viável ao método de peneiramento que ofereça análise granulométrica em tempo real e com o mínimo de interferência na linha de produção."

1.1 Contextualização do problema

Esta proposta insere-se no âmbito do programa MAI/DAI de Doutorado Acadêmico Industrial do CNPq, que tem como objetivo o desenvolvimento de pesquisas aplicadas à realidade e às necessidades do setor industrial brasileiro. De forma mais específica, a proposta se contextualiza na indústria de fertilizantes, com foco no desenvolvimento de tecnologias para a análise dimensional de grãos de adubo.

A análise granulométrica é um processo laboratorial essencial para garantir um nível adequado de qualidade na produção de fertilizantes, além de contribuir para a redução dos custos de produção. Esse processo utiliza peneiras com telas de diferentes tramas, semelhantes às apresentadas na Figura 2. As peneiras possuem aberturas variadas, conforme mostrado na Tabela 1. Os grãos analisados neste estudo possuem diâmetros que variam de 1 mm, sendo retidos na peneira de tela #18, até 4,75 mm, sendo retidos na peneira #5 e passando pela peneira #4, conforme indicado na Tabela 1.

Em condições ideais, o processo de produção de adubo opera dentro de faixas granulométricas entre 3,35 mm e 2,3 mm de diâmetro dos grãos.

Tabela 1: Peneiras utilizadas na indústria de fertilizantes.

Peneira	Abertura
#4	4.76 mm
#5	4.00 mm
#6	3.36 mm
#7	2.80 mm
#8	2.36 mm
#10	2.00 mm
#12	1.70 mm
#18	1.00 mm

Grãos que estiverem fora dessa faixa devem ser pulverizados e realimentados no início da linha de produção para serem novamente granularizados.

Atualmente, existem duas principais abordagens para realizar esse controle, descritas a seguir.

1.1.1 Controle de Granulometria por peneiramento

O controle de granulometria por peneiramento utiliza peneiras e um processo mecânico para sua execução, necessitando de interferência humana para seu funcionamento.

A Tabela 1 apresenta as categorias de peneiras utilizadas na indústria para o processo de peneiramento (*sieving*) semelhantes às peneiras ilustradas na Figura 2. Após a coleta e pesagem da amostra de fertilizante na linha de produção, ela é posicionada na peneira #4, que possui a maior abertura. As peneiras são empilhadas na mesma ordem da Tabela 1, com o tamanho das aberturas das telas decrescendo de cima para baixo, e devidamente posicionadas em uma máquina agitadora (Figura 3). A amostra então é agitada, realizando o processo de peneiramento dos grãos.

Em seguida, o conteúdo retido em cada peneira é novamente pesado, permitindo o cálculo da porcentagem presente em cada categoria analisada. Com base nesses dados, a fábrica pode calibrar sua produção.



Figura 2: Peneiras.

Na indústria de fertilizantes utilizam-se oito peneiras para classificação, qualquer retenção na peneira #6 e acima é considerada como grão grosso, e qualquer retenção abaixo da peneira #8 é considerada como grão fino. Os grãos retidos nas peneiras #7 e #8 são considerados de tamanho ideal, porém, os demais precisam ser decompostos em pó e retroalimentados à linha de produção, implicando em consumo de tempo e recursos.

O processo de peneiramento traz como vantagens a acurácia de medição, baixo custo de implementação, manuseio simples e disponibilidade rápida de resultado. Por outro lado, o método requer a coleta do produto, as peneiras podem sofrer danos, apresentam dificuldade na limpeza e, por não ser *on-line* existe um intervalo entre a coleta do produto e a análise, podendo apresentar disparidade com a realidade atual da linha de produção.



Figura 3: Agitador de peneiras.

1.1.2 Controle de granulometria por imagens

Como alternativa ao peneiramento, o controle de granulometria por imagens pode apresentar uma solução com maior efetividade. Para realizar a análise granulométrica utilizando visão computacional, as seguintes etapas devem ser executadas: na imagem, grãos devem ser separados em duas categorias, grãos e fundo, para permitir sua análise. Após o isolamento dos grãos, as dimensões dos grãos devem ser coletadas, analisando seu diâmetro, simulando o peneiramento mecânico, separando os mesmos a partir do seu diâmetro em grupos, retornando uma % de grãos presentes em cada categoria.

Entretanto a análise granulométrica por visão traz alguns desafios: o pequeno diâmetro dos grãos, somado a seu formato irregular, a homogeneidade das cores e a ausência de fundo acabam inviabilizando métodos mais simples de segmentação, dificultando a definição das suas bordas. Além disso, existe um alto índice de sobreposição entre os grãos que, ao serem segregados devido à vibração das esteiras, não representam a granulometria média da amostra. A vibração nas esteiras acarreta o deslocamento dos grãos e os mais grossos acabam localizando-se na superfície da amostra, enquanto grãos mais finos são movidos para a parte inferior.

Em adição aos problemas relacionados à geometria, à textura e à disposição dos grãos, somam-se os desafios relacionados ao cenário de operação. O chão de fábrica neste tipo de processo adiciona um quantitativo de ruídos na imagem, sejam eles:

- Presença de turbidez: a poeira presente na planta pode interferir no processamento de imagens, especialmente ao obstruir parcialmente a visão dos grãos, o que pode aumentar a dificuldade de sua segmentação;

- Borrão por movimento: causado pela movimentação da esteira, esse fenômeno pode

distorcer o formato do grão adquirido na imagem, dificultando sua análise granulométrica por visão computacional; e

- Baixo contraste devido à iluminação: a incidência de iluminação está diretamente relacionada à qualidade das imagens. Um alto nível de iluminação pode prejudicar a obtenção das imagens devido ao reflexo da poeira presente na fábrica, enquanto um nível de iluminação baixo degrada a informação, podendo acarretar a necessidade de um maior tempo de abertura do obturador da câmera, o que intensifica os borrões e dificulta seu tratamento.

Tecnologias atuais que utilizam o controle granulométrico por visão computacional enfrentam esses desafios desviando amostras de grãos do processo e analisando-as em um ambiente controlado à parte. Por exemplo, o sistema analisador de partículas CAMSIZER (Figura 4) utiliza um desvio da linha de produção para segregar algumas amostras, que são individualmente imageadas e processadas.

Essa abordagem utiliza uma série de desvios para produzir um fluxo esparso de grãos sobre um fundo com alto contraste de cor em relação aos fertilizantes, permitindo sua fácil segmentação e análise granulométrica.

Com esse tipo de maquinário, é possível obter uma leitura em tempo real da linha de produção, permitindo que a fábrica seja calibrada automaticamente, sem a necessidade de interferência humana, com base na produção atual.

Entretanto, o uso dessa tecnologia apresenta algumas desvantagens em comparação ao controle granulométrico por peneiramento. Dentre elas, destacam-se:

- análise limitada às amostras de um conjunto restrito de grãos desviados da linha de produção, não permitindo uma avaliação mais abrangente de todo o volume produzido;

- alto custo associado à implementação do maquinário especializado e sua adaptação dentro da linha de produção;

- necessidade da implementação de desvios na linha de produção para alimentar o maquinário de análise de partículas;

- devido à natureza química dos fertilizantes e sua abrasividade, o ambiente é extremamente hostil para qualquer tipo de maquinário, exigindo reposição frequente de peças e manutenção constante; e

- entre 5% e 1% do fertilizante produzido encontra-se na forma de pó, independentemente da calibração da fábrica. Somado à umidade gerada pela queima de cavacos para aquecimento e secagem do fertilizante, esse fator cria o ambiente ideal para a



Figura 4: CAMSIZER para npk e outros granulares [Microtrac 2025]

formação de camadas de fertilizante nas superfícies de qualquer maquinário. Devido a esse efeito, fábricas de fertilizantes devem realizar paradas programadas para remover o excesso de fertilizante depositado nas paredes do maquinário. Essa característica, por consequência, leva ao entupimento dos dispositivos, aumentando ainda mais a necessidade de manutenção.

1.2 Proposta deste trabalho

Como alternativa ao uso de desvios na linha de produção e ao alto investimento em maquinário especializado no âmbito do edital MAI/DAI CNPq, este trabalho explora a

possibilidade de aplicar visão computacional para estimar a granularidade dos fertilizantes por meio de câmeras distribuídas ao longo de toda a produção. As câmeras seriam instaladas diretamente acima das esteiras ao longo da planta, capturando em tempo real vistas superiores dos perfis de grãos.

O processo de análise granulométrica por visão computacional proposto neste trabalho baseia-se em imagens adquiridas a partir de uma vista superior da linha de transporte da fábrica. A proposta envolve a segmentação dos grãos presentes na imagem, conforme ilustrado na Figura 20, e a geração de uma imagem segmentada semelhante à Figura 21. Após a segmentação, os grãos são classificados em diferentes grupos com base no seu diâmetro, permitindo uma análise precisa e em tempo real da produção.

A implementação dessa abordagem pode trazer benefícios como um baixo custo de instalação, exigindo apenas investimento no mecanismo de captura de imagens, na rede de comunicação e em uma câmara de pressão positiva para reduzir o desgaste dos equipamentos. Não há necessidade de desvios na linha de produção, bastando apenas uma estrutura de suporte acima das esteiras para a instalação do mecanismo de captura das imagens.

Com um sistema *on-line*, torna-se possível obter uma leitura instantânea da produção, possibilitando o monitoramento da granulometria do processo ao longo de toda a linha de produção sem a interferência humana.

Entretanto, o passo de segmentação dos grãos representa o maior obstáculo deste trabalho.

1.3 Motivação

Problemas de segmentação figuram entre os principais obstáculos para a computação visual [Norozpour e Momenzadeh 2021] [Wu et al. 2017]. Neste trabalho, propõe-se explorar métodos de segmentação clássicos e de aprendizado de máquina para tratar imagens "densas e ruidosas", nas quais os objetos se encontram sobrepostos e aglomerados, e o imageamento apresenta ruídos inerentes devido à presença de poeira e vibração dos grãos. A maioria das técnicas clássicas de segmentação de imagens pode ser agrupada em três grandes categorias:

1. Técnicas baseadas em limite (*threshold*);
2. Técnicas baseadas em detecção de bordas; e
3. Técnicas baseadas em detecção de região [Abdulateef e Salman 2021].

Os métodos de segmentação baseados em *threshold* estão entre as técnicas mais simples para realizar a segmentação de objetos, baseando sua segmentação apenas na intensidade dos *pixel*, ignorando as características espaciais da imagem. Possuem seu funcionamento baseado na ideia de separar uma imagem em duas regiões, região frontal e

região de fundo, utilizando a diferença de intensidade para realizar a separação. Partes da imagem que possuem valores inferiores a um determinado valor de *threshold* pertencem a um determinado grupo. Devido a este funcionamento, a definição de um valor de *threshold* se torna chave para realização de uma boa segmentação. Entretanto, uma alta homogeneidade de intensidades na imagem pode dificultar a escolha de um valor ideal de *threshold*.

Os métodos de segmentação baseados na detecção de bordas apoiam-se na premissa de que cada objeto na imagem é cercado por uma descontinuidade na intensidade dos *pixels*. Para realçar essas bordas, são aplicados diversos filtros que evidenciam variações bruscas de intensidade, classificando as regiões resultantes como borda ou não borda. Em seguida, é possível utilizar técnicas de pós-processamento, como a segmentação por limiar *threshold*, para concluir a separação entre as bordas e o fundo, finalizando a identificação dos objetos. Embora essa abordagem possa parecer promissora para o problema de segmentação em estudo, métodos determinísticos clássicos baseados em detecção de bordas tendem a perder efetividade em imagens complexas e ruidosas, sobretudo quando o número de objetos presentes é elevado [Abdulateef e Salman 2021].

Por sua vez, os métodos de segmentação baseados em região partem do princípio de que os *pixels* de uma determinada região compartilham valores similares. Essas técnicas podem ser divididas, em linhas gerais, em duas principais abordagens: Crescimento de regiões e Divisão de regiões e fusão. O método de crescimento de regiões inicia-se com um conjunto de sementes que, em tese, representam as diferentes regiões. Em seguida, essas regiões são ampliadas ao incorporar *pixels* adjacentes, tendo como base um critério de similaridade pré-definido. Esse processo se repete até que nenhum pixel vizinho atenda mais ao critério estabelecido. Como o número de regiões depende diretamente da quantidade de sementes inicializadas, a calibração adequada desse método constitui um grande desafio [Adams e Bischof 1994].

Por sua vez, o método de divisão e fusão de regiões adota uma abordagem iterativa baseada na segmentação por regiões. Inicialmente, as regiões adjacentes são comparadas e, caso apresentem homogeneidade, são fundidas. No entanto, em virtude de sua característica de fundir regiões homogêneas, essa técnica encontra consideráveis dificuldades para segmentar os objetos analisados neste estudo [14].

Dadas as limitações das abordagens clássicas mencionadas, é improvável que uma única calibração desses métodos seja suficiente para alcançar resultados satisfatórios. Além da motivação voltada para a solução de um problema complexo no processo de fabricação de adubo, este trabalho também busca enfrentar os desafios inerentes à segmentação visual de objetos em imagens densas e ruidosas. Entre esses desafios, destacam-se a necessidade de segmentar um grande número de objetos em uma única imagem, a presença de bordas difusas, a sobreposição entre elementos e a similaridade na composição dos grãos. Além disso, fatores típicos do ambiente industrial, como ruídos

de movimento, variações de iluminação e a presença de poeira, tornam o problema ainda mais desafiador. Diante dessas características, a segmentação visual precisa e eficiente desses elementos em cenário tão adverso se configura como uma tarefa computacional complexa, exigindo abordagens que vão além dos métodos convencionais.

Uma abordagem promissora para lidar com os desafios de segmentação e calibração é o uso de aprendizado de máquina, especialmente redes neurais profundas, para a segmentação dos grãos. Inspiradas no funcionamento dos neurônios biológicos, essas redes processam informações propagando sinais entre neurônios conectados em múltiplas camadas. A partir de dados de entrada, a rede codifica e transmite os sinais por suas camadas até a saída, onde os resultados são comparados com valores esperados. Durante o treinamento, os pesos das conexões entre neurônios são ajustados com base em uma função de aprendizado pré-definida, minimizando a diferença entre a saída gerada e a resposta esperada. Esse processo iterativo permite que a rede aprenda padrões complexos nos dados, tornando-se uma solução mais adaptável e eficaz para os desafios da segmentação de grãos em imagens [Agatonovic-Kustrin e Beresford 2000].

Nos últimos anos, o uso de redes neurais tem se expandido significativamente, impulsionando avanços expressivos nas diversas áreas em que são aplicadas. Esse crescimento está diretamente associado à evolução do hardware, que possibilita a execução de cálculos complexos de forma eficiente, além do aumento da disponibilidade de dados anotados, um fator essencial para aprimorar a qualidade do aprendizado das redes. O acesso a grandes volumes de dados e a melhoria na capacidade computacional têm viabilizado o desenvolvimento de modelos mais precisos e robustos, ampliando ainda mais o impacto do aprendizado de máquina em aplicações práticas [Ghosh et al. 2019].

A aplicação de redes neurais tem se destacado pela notável capacidade de generalização e adaptação ao aprendizado, frequentemente proporcionando resultados mais eficazes na solução de problemas complexos. Métodos baseados em aprendizado de máquina requerem, inicialmente, um grande volume de dados anotados para treinamento. No entanto, uma vez treinados, esses modelos podem ser aplicados a novos dados semelhantes sem a necessidade de anotações adicionais, tornando o processo mais eficiente e escalável.

Neste estudo será averiguado se uma abordagem baseada em visão computacional e aprendizado de máquina pode ser uma solução eficaz para a segmentação de objetos com sobreposição e aglomeração. A capacidade desses modelos de identificar padrões sutis e adaptar-se a diferentes condições poderá possibilitar uma segmentação mais precisa e robusta, superando as limitações dos métodos clássicos.

Além da falta de dados anotados em escala, o cenário físico em si desempenha um papel fundamental na complexidade do problema. Em ambientes densos de grãos, ocorrem ruídos consideráveis, iluminação desigual e sobreposições mais intensas, tornando o processo de segmentação ainda mais desafiador. Quando o ambiente é intrinsecamente

ruidoso e heterogêneo, torna-se imprescindível não apenas reunir um conjunto diversificado de amostras para treinamento, mas também garantir que essas anotações reflitam fielmente as condições reais. Dessa forma, o desafio de escassez de dados anotados e as peculiaridades do ambiente ruidoso são fatores igualmente críticos para o sucesso de abordagens baseadas em aprendizado de máquina.

1.4 Hipótese Científica

A hipótese científica deste trabalho baseia-se na premissa de que métodos computacionais para segmentação de imagens densas e ruidosas, especialmente aqueles fundamentados em aprendizado de máquina, possuem uma capacidade superior de adaptação e generalização em comparação com os métodos clássicos. Essa vantagem possibilita a inferência da granularidade dos grãos em imagens capturadas do topo das esteiras em linhas de produção de fertilizantes, desde que houvesse um adequado ajuste do modelo por meio de aprendizado supervisionado.

Por outro lado, utilização combinada de imagens reais com aquelas segmentadas por métodos clássicos pode aprimorar a precisão do processo de segmentação. Embora esses métodos determinísticos, isoladamente, não sejam suficientes para lidar com imagens complexas, eles fornecem informações estruturadas que enriquecem o treinamento de redes neurais. Como resultado, uma abordagem híbrida, envolvendo aprendizado de máquina e métodos clássicos de segmentação, poderia aumentar a capacidade do modelo de lidar com objetos complexos e sobrepostos, reduzindo ambiguidades e tornando a segmentação mais robusta e eficiente.

Também, o uso combinado de dados sintéticos e reais poderia expandir o espaço de variação do conjunto de treinamento, aperfeiçoando a capacidade de generalização do modelo e viabilizando a calibração de um número maior de parâmetros; dessa forma, tornando possível lidar melhor com a alta escalabilidade de configurações necessárias à segmentação de imagens densas e complexas, garantindo maior robustez ao sistema em contextos variados.

1.5 Objetivos

O objetivo principal desta tese é propor um modelo híbrido baseado em aprendizado de máquina e métodos clássicos de visão computacional para segmentar grãos em imagens de topo obtidas em ambiente industrial.

O modelo deverá tratar os desafios inerentes à análise granulométrica visual de fertilizantes em ambientes industriais que envolvem tanto a sobreposição densa dos grãos e ruídos presentes nas imagens, como a dificuldade de se obter *datasets* anotados para seus ajustes.

1.5.1 Objetivos Específicos

- Identificar e analisar as principais características das imagens capturadas do topo das esteiras no chão de fábrica, considerando aspectos como iluminação, ruído e sobreposição de grãos;

- Implementar e avaliar métodos clássicos e do estado da arte em segmentação, comparando suas vantagens e limitações no contexto da análise granulométrica de fertilizantes;

- Propor um método para criação e uso de *datasets* anotados capaz de permitir um ajuste adequado do modelo com base na geração e anotação de imagens sintéticas, reais em ambiente controlado e reais em ambiente real: e

- Propor um modelo híbrido envolvendo aprendizado de máquina e algoritmos clássicos de segmentação, capaz de lidar com os desafios inerentes ao ambiente industrial;

1.6 Publicações

Devido às restrições encontradas durante o desenvolvimento inicial, foi criado um simulador capaz de gerar imagens virtuais de superfícies totalmente cobertas por grãos de fertilizantes, por meio da colagem de imagens individuais. Esse desenvolvimento resultou na primeira publicação:

- **Simulated Dataset to Verify the Overlapping and Segregation Problem on Computer Vision Granullometry of Fertilizers** [Cezar et al. 2020], que aborda o uso do simulador para gerar datasets virtuais, permitindo analisar problemas relacionados à sobreposição e segregação de grãos.

Posteriormente, os estudos foram aprofundados utilizando métodos determinísticos, como os operadores Canny e Otsu, com a adição de ruídos artificiais – poeira, variação de iluminação e movimento – para avaliar a robustez das técnicas de segmentação. Esse estudo gerou a publicação:

- **Grain Surface Simulator to Averiguate the Overlapping and Noise Problems on Computer Vision Granullometry of Fertilizers** [Goulart et al. 2020], que enfatiza os desafios na obtenção de dados anotados e a importância da simulação de imagens realistas com grãos sobrepostos.

Em 2021, explorou-se a aplicação de métodos convencionais em imagens reais, investigando como diferentes tipos de ruídos afetam a segmentação, resultando no artigo:

- **VBM Granulometry for Real Applications in the Fertilizer Industry** [Traversi et al. 2021], que demonstra a viabilidade desses métodos em contextos industriais reais, destacando a influência negativa dos ruídos, como poeira e iluminação, sobre a qualidade da segmentação.

Diante das limitações das abordagens determinísticas, passou-se a investigar métodos [Traversi et al. 2021] fundamentados em inteligência artificial, começando pela aplicação da rede U-Net. Esse esforço resultou na publicação:

- **A Neural Network for Segmentation of Fertilizer Grain with Multiple Sizes and Without Background** [Traversi et al. 2022], destacando o potencial de técnicas de aprendizado profundo para uma segmentação mais precisa e robusta.

Após a consolidação do estudo com U-Net, tornou-se evidente a necessidade de *datasets* mais representativos. Foram então desenvolvidos três conjuntos distintos:

- **Dataset Virtual**: gerado pelo simulador com anotações perfeitas. No entanto, devido ao reaproveitamento frequente dos mesmos grãos para compor as imagens, apresenta uma tendência ao *overfitting*.
- **Dataset em Ambiente Controlado**: imagens obtidas sob condições ideais e segmentadas manualmente, oferecendo maior diversidade. Contudo, a anotação manual introduziu desafios de precisão.
- **Dataset em Ambiente Relevante**: obtido diretamente da fábrica, apresentando ruídos significativos relacionados à poeira e movimentação dos grãos. A complexidade na segmentação manual limitou o volume de imagens disponíveis para testes.

Esses três *datasets* foram detalhadamente apresentados na publicação:

- **Development of Comprehensive Fertilizer Datasets: Enhancing Precision Agriculture through Data-Driven Insights** [Traversi et al. 2024], descrevendo sua criação, desafios e potenciais aplicações na agricultura de precisão.

1.7 Organização da Tese

Esta tese está organizada de forma a conduzir o leitor de maneira lógica e progressiva pelos tópicos abordados. A seguir, apresenta-se um breve resumo do conteúdo de cada capítulo:

- **Capítulo 1:** Apresenta a contextualização do problema, abordando a motivação que impulsionou a pesquisa, a hipótese científica e os objetivos do estudo.

- **Capítulo 2:** Desenvolve o referencial teórico, onde são explicados os principais conceitos aplicados nesta tese, os métodos de segmentação, as métricas de avaliação e o estado da arte relacionado à segmentação.
- **Capítulo 3:** Descreve a metodologia utilizada para a obtenção de *dataset* de imagens densas com sobreposição, assim como uma análise comparativa da eficácia entre conjuntos de dados sintéticos, controlados e reais, e uma avaliação da efetividade considerando diferentes proporções de cada *dataset*.
- **Capítulo 4:** Avalia diversas redes CNN com o objetivo de identificar a melhor arquitetura para servir de *encoder* na rede proposta neste trabalho.
- **Capítulo 5:** Apresenta o modelo proposto nesta tese, chamado de GranSeg e detalha sua estrutura e a inicialização dos parâmetros.
- **Capítulo 6:** Exibe os resultados obtidos pelo método GranSeg, comparando-os com redes U-net e suas variações, e realiza uma comparação com o método SAM, que representa o estado da arte.
- **Capítulo 7:** Conclui o trabalho com uma síntese dos principais resultados e uma discussão das implicações dos achados.

2 REFERENCIAL TEÓRICO

Como alternativa ao método clássico de peneiramento mecânico, acredita-se que técnicas de visão computacional possam ser empregadas para essa finalidade, oferecendo uma abordagem mais eficiente e automatizada. O Processamento Digital de Imagens (PDI) consiste em um conjunto de métodos voltados à manipulação e análise digital de imagens por meio de algoritmos. Esses métodos podem ser determinísticos, garantindo resultados consistentes para uma mesma configuração, ou não determinísticos, como ocorre com redes neurais, cujo desempenho pode variar devido ao processo de aprendizado e à convergência do treinamento [56].

Para a análise das imagens, a identificação dos limites dos objetos é um passo essencial, pois, em muitos casos, esses limites permitem a distinção entre os elementos de interesse e o fundo. Dessa forma, algoritmos de PDI podem ser aplicados para a segmentação e posterior análise do tamanho das partículas.

Neste contexto, serão apresentados os principais conceitos relacionados à segmentação de imagens, explorando as duas abordagens utilizadas neste estudo: métodos clássicos e redes neurais. Ambos serão avaliados quanto à sua eficácia na segmentação de grãos ou minerais, considerando os desafios impostos por imagens com objetos sobrepostos.

Além disso, serão discutidas métricas utilizadas para a avaliação da segmentação, permitindo a comparação dos resultados e a validação dos métodos propostos.

Por fim, serão demonstradas técnicas de *data augmentation*, utilizadas para expandir os conjuntos de dados anotados, aumentando a diversidade e a robustez dos modelos de segmentação, contribuindo para uma melhor adaptação a diferentes cenários.

2.1 Principais Conceitos

Busca-se neste trabalho desenvolver um método efetivo de segmentação baseado em PDI, capaz de permitir a estimativa de granulometria a partir de imagens. A seguir, descrevem-se alguns conceitos associados ao processamento de imagens que são utilizados neste trabalho.

2.1.1 Visão Computacional

A visão computacional é um campo interdisciplinar que visa interpretar e extrair informações significativas a partir de imagens digitais, empregando métodos algorítmicos avançados para reproduzir, e até mesmo superar, capacidades visuais humanas. As técnicas envolvidas permitem identificar, classificar e analisar padrões visuais complexos presentes em imagens ou vídeos, contribuindo para áreas como reconhecimento facial, veículos autônomos e diagnóstico médico.

2.1.2 Redes Neurais Convolucionais

Redes neurais convolucionais (CNNs, do inglês *Convolutional Neural Networks*) são estruturas especializadas para o processamento de imagens, compostas por múltiplas camadas que aprendem automaticamente filtros visuais específicos. Esses filtros identificam características importantes como bordas, texturas e formas por meio da aplicação de convoluções, transformando os dados de entrada em representações intermediárias cada vez mais abstratas e refinadas, facilitando a classificação e reconhecimento visual.

2.1.3 Janela Deslizante (Sliding Window)

A técnica da janela deslizante (*sliding window*) é uma abordagem que permite analisar sistematicamente diferentes regiões da imagem, movendo uma janela fixa através de toda a extensão da imagem para realizar a classificação ou detecção local. Esse método é amplamente utilizado em conjunto com redes convolucionais para identificar múltiplos objetos ou padrões dentro de uma única imagem, permitindo que a rede neural concentre sua atenção em áreas específicas e relevantes.

2.1.4 Camada Convolucional

A camada convolucional é um dos componentes fundamentais de redes neurais convolucionais (CNNs). Ela opera aplicando filtros (ou kernels) sobre uma imagem de entrada para extrair características relevantes. Matematicamente, a operação de convolução pode ser expressa como:

$$(I * K)(x, y) = \sum_i \sum_j I(i, j)K(x - i, y - j), \tag{1}$$

$$Z(x, y) = (i * k)(x, y) + b, A(x, y) = f(Z(x, y)), \tag{2}$$

onde I representa a imagem de entrada e K é o kernel. Essa operação realça padrões específicos, como bordas, texturas e formas.

2.1.5 Camada de Pooling e Downsampling

A camada de pooling é usada para reduzir a dimensionalidade da imagem de entrada, preservando informações essenciais e tornando a rede mais eficiente. Os dois tipos principais são:

- **Max Pooling**: Seleciona o maior valor dentro de uma janela deslizante.
- **Average Pooling**: Calcula a média dos valores na janela.

Essas operações realizam um downsampling, reduzindo a resolução espacial da imagem e tornando o modelo mais robusto a variações na entrada.

2.1.6 Operações de Upsampling

O upsampling é utilizado para recuperar a resolução espacial perdida no downsampling, sendo essencial para redes como a U-Net. Os principais métodos incluem:

2.1.6.1 Interpolação Bilinear

A interpolação bilinear é uma técnica matemática frequentemente usada no processamento de imagens para estimar o valor de intensidade em coordenadas não inteiras, considerando os quatro pixels vizinhos mais próximos. Seja um ponto de interesse (x, y) em coordenadas contínuas, onde desejamos calcular a intensidade $I(x, y)$. Suponha que os valores conhecidos dos pixels vizinhos mais próximos estejam localizados nos pontos inteiros (x_0, y_0), (x_1, y_0), (x_0, y_1) e (x_1, y_1), com intensidades respectivamente dadas por I_{00}, I_{10}, I_{01} e I_{11}. Definindo ainda os pesos relativos como $\alpha = x - x_0$ e $\beta = y - y_0$, a interpolação bilinear calcula o valor estimado pela média ponderada das intensidades vizinhas da seguinte forma:

$$I(x, y) = (1 - \alpha)(1 - \beta)I_{00} \ + \ \alpha(1 - \beta)I_{10} \ + \ (1 - \alpha)\beta I_{01} \ + \ \alpha\beta I_{11} \qquad (3)$$

Dessa maneira, α e β controlam a influência relativa dos quatro pixels vizinhos, garantindo uma transição suave e contínua entre eles. Esse método é especialmente útil para o redimensionamento, rotação e transformação geométrica de imagens, onde se deseja preservar a qualidade visual da imagem original.

2.1.6.2 Convoluções Transpostas

As convoluções transpostas aprendem a reconstruir a informação perdida aplicando filtros treináveis de forma inversa às convoluções normais. Elas expandem a entrada usando preenchimento de zeros e aplicam uma convolução regular.

2.1.6.3 Unpooling

A operação de *unpooling* busca reverter os efeitos da redução dimensional provocada pelo processo de *max pooling*. Enquanto o *max pooling* reduz a resolução espacial da imagem ou do mapa de características ao preservar apenas o valor máximo dentro de cada janela local, o *unpooling* realiza o procedimento inverso, objetivando reconstruir as dimensões espaciais originais. Uma abordagem amplamente utilizada consiste em armazenar as posições exatas dos elementos máximos durante a etapa de *max pooling*, criando um conjunto de índices ou máscaras. Posteriormente, na etapa de *unpooling*, esses índices são empregados para recolocar cada valor comprimido precisamente na posição original que ocupava antes da redução, preenchendo as posições restantes com zeros ou interpolando valores intermediários. Tal procedimento permite uma reconstrução mais precisa da resolução original, garantindo a manutenção de informações espaciais importantes para aplicações como segmentação semântica ou autoencoders convolucionais.

2.1.7 *Thresholding*

A técnica de *thresholding* consiste em segmentar uma imagem separando suas regiões em duas classes distintas com base em um valor de limiar pré-definido. Esse limiar é normalmente determinado pelo usuário e aplicado aos valores de intensidade luminosa da imagem.

Dada uma imagem $I(x, y)$, a segmentação por limiarização gera uma imagem binária $b(x, y)$ ao aplicar um limiar T, conforme a seguinte regra:

$$b(x,y) = \begin{cases} 0, & \text{se } I(x,y) \leq T \\ 255, & \text{se } I(x,y) > T \end{cases} \tag{4}$$

Normalmente, os *pixels* de interesse (objetos) são marcados como 0 (preto) e os *pixels* do fundo como 255 (branco). Essa técnica é amplamente utilizada no pré-processamento de imagens para destacar elementos de interesse e facilitar etapas subsequentes de segmentação e análise [Melo et al. 2014].

2.1.8 Equalização de Histograma Adaptativo

A equalização de histograma é uma técnica que redistribui os níveis de intensidade de uma imagem para obter uma distribuição mais homogênea dos tons de cinza, melhorando o contraste. No entanto, em algumas situações, podem existir regiões da imagem com diferentes níveis de contraste, o que pode resultar em extremos de brilho ou escuridão quando a equalização é aplicada de forma global.

Para lidar com esses casos, a equalização de histograma adaptativo é uma abordagem mais eficiente. Essa técnica divide a imagem em regiões menores de tamanho semelhante e aplica a equalização de histograma separadamente em cada uma dessas seções. Dessa

forma, a distribuição de intensidade é ajustada localmente, permitindo uma melhoria mais consistente do contraste em diferentes áreas da imagem [32].

2.1.9 Segmentação em Visão Computacional

A segmentação em visão computacional é uma etapa fundamental que permite dividir uma imagem em regiões significativas, facilitando a análise e a interpretação dos dados visuais. Nesta subseção, são abordadas três técnicas principais de segmentação: semântica, de instância e panóptica.

Segmentação Semântica: Nesta abordagem, cada pixel da imagem é classificado em uma categoria predefinida, como (carro, pessoa ou árvore). Todos os pixels pertencentes à mesma classe recebem o mesmo rótulo, proporcionando uma visão global da cena. Contudo, essa técnica não diferencia objetos distintos da mesma classe; por exemplo, todos os carros são rotulados igualmente como (carro).

Segmentação de Instância: Essa técnica vai além da semântica, identificando e separando cada objeto individualmente mesmo que pertença à mesma categoria. Assim, cada objeto recebe uma máscara e uma identificação única, o que é essencial para aplicações que exigem a contagem e a análise detalhada dos elementos presentes na imagem. Essa abordagem, no entanto, demanda maior complexidade computacional.

Segmentação Panóptica: A segmentação panóptica integra as vantagens das abordagens semântica e de instância. Nela, cada pixel recebe uma etiqueta de classe e, quando aplicável, uma identificação individual para objetos distintos. Essa técnica proporciona uma análise abrangente, combinando a visão global da cena com a diferenciação dos objetos, sendo particularmente útil em aplicações complexas, como a percepção para veículos autônomos.

Considerações sobre Técnicas de Segmentação: Cada técnica de segmentação apresenta vantagens e limitações específicas. A segmentação semântica oferece uma visão global da cena ao atribuir o mesmo rótulo a todos os pixels de uma mesma classe, enquanto a segmentação de instância destaca individualmente cada objeto, possibilitando uma análise detalhada. Em contraste, a segmentação panóptica integra os pontos fortes dessas abordagens, fornecendo uma análise completa e integrada da imagem. Neste trabalho, aplicamos a segmentação de instância para distinguir e segmentar grãos semelhantes.

2.1.10 Redes Neurais Convolucionais

Assim como outras redes neurais, as Redes Neurais Convolucionais (CNNs, do inglês *Convolutional Neural Networks*) são baseadas no funcionamento dos neurônios artificiais, organizados em camadas. Os neurônios entre as camadas são conectados por pesos. A primeira camada da rede é chamada de camada de entrada, enquanto a última é a camada de saída. Todas as camadas situadas entre a entrada e a saída são conhecidas como *hidden*

layers (camadas ocultas).

As CNNs possuem, pelo menos, uma camada convolucional como *hidden layer* [Kattenborn et al. 2021]. Para melhorar a capacidade de aprendizado, as redes CNN não processam diretamente toda a imagem de uma só vez. Em vez disso, utilizam um filtro conhecido como *patch* ou *kernel*, que possui um tamanho menor que a imagem original e percorre a imagem extraindo informações relevantes.

Entre as camadas pertencentes às redes neurais convolucionais (CNNs), destacam-se as camadas de Max Pooling, responsáveis por reduzir a dimensionalidade do mapa de características extraído pelas camadas convolucionais anteriores. Esse procedimento é realizado selecionando o maior valor presente em cada região específica (*patch*) do mapa de características. Tal operação resulta em uma versão reduzida e mais condensada do mapa original, o que auxilia não apenas na redução do custo computacional, como também no controle da sensibilidade da rede a pequenas variações na posição dos objetos. Ao preservar as características mais relevantes enquanto reduz as redundâncias, as camadas de Max Pooling contribuem significativamente para melhorar a eficiência e a robustez da rede neural.

Em contrapartida às camadas convolucionais e de pooling, as camadas totalmente conectadas (*fully connected*, FC) têm um papel fundamentalmente distinto. Elas recebem integralmente as informações geradas nas etapas anteriores da rede, sem a utilização de filtros ou operações de redução dimensional adicionais. Nessas camadas, cada neurônio é conectado diretamente a todos os neurônios da camada anterior, possibilitando uma interpretação global e abrangente das características extraídas. A classificação final ou tomada de decisão da rede neural é, portanto, realizada com base em uma combinação ponderada dessas informações, utilizando pesos aprendidos ao longo do treinamento. Em geral, as camadas totalmente conectadas encontram-se no estágio final das redes CNN, sendo responsáveis por produzir o resultado conclusivo e definir a classe ou categoria da imagem analisada. Tal organização estrutural permite que a rede realize uma classificação robusta e acurada a partir das informações previamente extraídas pelas camadas convolucionais e de pooling.

2.1.11 Funcionamento das Camadas Neurais

Para a transmissão e processamento de informações dentro de uma rede neural, as camadas recebem dados de entrada sob a forma de valores numéricos, representando as características extraídas ou os resultados intermediários das camadas anteriores. Cada valor numérico que chega até um determinado neurônio é então multiplicado por um peso específico, previamente ajustado durante o treinamento da rede. Esses pesos podem assumir valores positivos ou negativos, refletindo a influência e a relevância que cada característica ou informação terá na composição da saída do neurônio. Após essa etapa inicial de multiplicação, realiza-se uma soma cumulativa dos valores obtidos após serem

multiplicados pelos respectivos pesos.

Além dos pesos, cada neurônio possui um termo adicional chamado *bias*. O *bias* é um valor escalar que permite à rede neural deslocar a função de ativação, tornando o modelo mais flexível e garantindo que a ativação do neurônio possa ocorrer mesmo quando todas as entradas forem iguais a zero. Assim, o resultado final da soma dos valores multiplicados pelos pesos recebe o acréscimo desse termo *bias*. O resultado dessa operação é então submetido a uma função de ativação. Essa função tem o papel crucial de definir o sinal a ser transmitido para a próxima camada da rede neural, além de introduzir não linearidade ao modelo. Tal não linearidade possibilita que a rede neural capture relações e padrões complexos nos dados, permitindo que tome decisões precisas e adaptativas diante das mais diversas situações ou conjuntos de entradas.

2.1.12 Funções de Ativação

As funções de ativação podem ser classificadas em dois tipos principais:

Funções lineares: transmitem o valor obtido diretamente, sem alterações. Esses valores podem ser positivos, negativos ou zero. Funções não lineares: modificam o comportamento da saída do neurônio de acordo com o valor do somatório ponderado.

Entre as funções de ativação não lineares mais utilizadas, destacam-se:

- Função Sigmoide: gera valores de saída entre 0 e 1. Quando o somatório ponderado possui valores negativos, a saída se aproxima de zero; quando o somatório é zero, a saída é aproximadamente 0,5; e para valores positivos, a saída se aproxima de 1.

- Função ReLU (*Rectified Linear Unit*): retorna zero para valores negativos ou iguais a zero e se comporta como uma função linear para valores positivos, facilitando o treinamento de redes profundas [51].

-

Essas funções desempenham um papel fundamental na capacidade de aprendizado da rede, permitindo que o modelo capture padrões complexos nos dados de entrada.

2.1.13 Dados Anotados

É a nomenclatura dada para dados considerados verdadeiros, em redes neurais são chamados de dados anotados [Kondermann 2013] os dados utilizados para o treinamento das redes contendo a resposta ideal esperada da rede.

2.1.14 Transfer Learning

Transfer Learning [67], ou aprendizado por transferência, se refere ao ato de utilizar um modelo pré-treinado de uma rede neural em um novo problema. Muitos problemas são similares em certos níveis, e redes neurais podem aproveitar o conhecimento obtido em treinos prévios para melhorar sua generalização em novos problemas. Nesse trabalho,

maior parte das redes foi previamente treinada no *dataset Imagenet*, um vasto conjunto de imagens classificadas. Utilizar esse conhecimento pré-treinado permite que redes neurais possam convergir mais rapidamente em um novo problema, utilizando um número de dados menor.

2.1.15 Skip Connections

Skip Connections, como o nome sugere, são conexões que pulam certas camadas, propagando a saída de uma camada da rede para camadas mais à frente, pulando uma ou mais camadas no caminho.

2.1.16 *Encoder* em Redes Neurais Convolucionais

O termo *encoder* refere-se ao processo de conversão de informações em um formato mais compacto ou adequado para processamento posterior. Em redes neurais convolucionais (CNNs), um *encoder* é composto por um conjunto de camadas responsáveis por transformar uma entrada de alta dimensão, como uma imagem, em uma representação mais simplificada e significativa.

Tipicamente, um *encoder* é formado por múltiplas camadas convolucionais, seguidas por camadas de pooling ou operações de downsampling, que reduzem gradualmente a resolução espacial da imagem, preservando as informações mais relevantes. Esse processo tem como objetivo extrair as características essenciais da entrada, filtrando detalhes redundantes e permitindo que a rede aprenda representações mais robustas.

A saída do *encoder* pode variar dependendo da aplicação. Em alguns casos, pode ser um mapa de características reduzido, contendo informações extraídas da imagem de entrada em uma forma mais condensada. Em outros, pode ser uma saída binária ou um vetor de características que representa a informação mais relevante da entrada original.

O *encoder* desempenha um papel fundamental em arquiteturas como *autoencoders*, redes U-Net e modelos de segmentação semântica, onde a compressão da informação facilita a reconstrução ou classificação posterior [8].

2.1.17 *Decoder*

O *decoder* executa o processo inverso do *encoder*, restaurando a informação comprimida para um formato mais próximo do original. Em redes neurais convolucionais, o *decoder* tem a função de aumentar a resolução dos mapas de características gerados pelo *encoder*, reconstruindo a estrutura espacial dos dados.

Esse processo geralmente é realizado por meio de técnicas de *upsampling*, como interpolações bilineares, convoluções transpostas (*deconvolutions*) ou operações de *unpooling*. No caso do *unpooling*, o *decoder* pode utilizar os índices armazenados durante as operações de *pooling* do *encoder* para reposicionar os valores nos locais corretos, restaurando melhor a estrutura da imagem original [8].

O *decoder* é amplamente utilizado em modelos de segmentação de imagens, *auto-encoders* e redes do tipo U-Net, onde a recuperação da estrutura espacial da imagem é essencial para garantir uma saída coerente e útil.

2.1.18 *Encoder-Decoder*

A arquitetura *Encoder-Decoder* é um modelo de rede neural projetado para transformar uma entrada de alta dimensão em uma representação comprimida por meio do *encoder*, e posteriormente restaurá-la utilizando o *decoder*. O *encoder* reduz progressivamente a dimensão espacial da entrada ao mesmo tempo que extrai suas características mais relevantes, enquanto o *decoder* reconstrói a saída de modo a preservar a estrutura e os detalhes essenciais da imagem.

Esse tipo de arquitetura é amplamente utilizado em tarefas como segmentação semântica, compressão de imagens, remoção de ruído, super-resolução e até mesmo tradução automática em modelos de aprendizado profundo [8].

Nas redes U-Net, por exemplo, a arquitetura *Encoder-Decoder* é complementada por conexões de atalho (*skip connections*), que ajudam a preservar detalhes espaciais ao longo do processo de reconstrução, minimizando a perda de informações.

2.1.19 *Autoencoder*

Os *autoencoders* [Kramer 1991] são redes neurais utilizadas principalmente para redução de ruído nos dados e para aprendizado de representações comprimidas. Eles funcionam removendo correlações redundantes entre as variáveis do problema, o que pode ser útil na redução de dimensionalidade e na extração de características relevantes.

O objetivo principal de um *autoencoder* é reconstruir sua entrada na saída, minimizando a diferença entre ambas. Para isso, a rede é treinada de forma não supervisionada, utilizando os próprios dados de entrada como rótulos. O modelo é composto por duas partes principais:

- *Encoder*: transforma a entrada em uma representação latente de menor dimensionalidade, preservando as informações mais relevantes. - *Decoder*: reconstrói a entrada original a partir da representação latente gerada pelo *encoder*.

Nos experimentos realizados neste trabalho, um *autoencoder* será aplicado utilizando a estrutura da rede U-Net. O objetivo é inicializar os pesos da U-Net com valores pré-treinados, permitindo que a rede aprenda uma representação inicial útil e, assim, facilite a convergência durante o treinamento. Essa estratégia pode melhorar a estabilidade e a eficiência do aprendizado, especialmente em tarefas de segmentação.

2.2 Segmentação

Diferentes abordagens para o problema de segmentação são amplamente exploradas na literatura, destacando-se os métodos tradicionais baseados em técnicas de Processamento Digital de Imagens (PDI) e abordagens mais recentes utilizando redes neurais artificiais.

Os métodos tradicionais de segmentação baseados em PDI podem ser classificados em três categorias principais.

2.2.1 Segmentação Baseada em Região

A segmentação baseada em região consiste em agrupar *pixels* com características semelhantes para formar regiões homogêneas dentro da imagem. Essa abordagem parte do princípio de que *pixels* pertencentes ao mesmo objeto compartilham propriedades comuns, como intensidade, cor ou textura, permitindo a separação dos elementos presentes na imagem com base nesses atributos.

Algumas das técnicas mais comuns incluem:

- *Watershed* [Sarkar et al. 2018], um método de segmentação morfológica que interpreta a imagem como uma superfície tridimensional, onde os níveis de intensidade representam elevações do terreno. Nesse modelo, áreas de baixa intensidade são tratadas como vales (*bacias hidrográficas*) e áreas de alta intensidade correspondem a picos. A segmentação ocorre simulando um processo de inundação: a água preenche os vales e, à medida que o nível sobe, barreiras são erguidas para impedir a fusão entre diferentes regiões. Esse método é eficaz na separação de objetos que apresentam contornos bem definidos, mas pode sofrer de supersegmentação, especialmente em imagens ruidosas, sendo comum o uso de técnicas auxiliares, como a aplicação de filtros ou o uso de marcadores.

- *K-means* [Liu e Yu 2009], um algoritmo de clusterização amplamente utilizado na segmentação de imagens. Ele opera dividindo os *pixels* em K grupos distintos, minimizando a variabilidade dentro de cada grupo. Inicialmente, K centróides são definidos aleatoriamente, e cada *pixel* é atribuído ao cluster mais próximo com base em sua distância em relação ao centróide. O processo é iterativamente refinado até que a variação dentro dos *clusters* seja minimizada. Embora eficiente, o *K-means* apresenta limitações quando a segmentação envolve objetos com distribuições complexas de intensidade ou quando o número K não é bem definido.

2.2.2 Segmentação Baseada na Detecção de Bordas

A segmentação baseada na detecção de bordas busca identificar os contornos dos objetos na imagem analisando mudanças abruptas na intensidade dos *pixels*. Essa aborda-

gem é especialmente útil quando os objetos possuem bordas bem definidas, permitindo a separação dos elementos da cena com base nas transições de intensidade.

Os principais métodos incluem:

- *Prewitt* [Rahmawati et al. 2021], um operador diferencial que calcula a variação da intensidade da imagem utilizando filtros convolucionais. Ele fornece uma estimativa do gradiente nos eixos horizontal e vertical, destacando áreas de transição brusca entre regiões. Apesar de ser um método simples e computacionalmente eficiente, é sensível a ruídos e pode apresentar limitações na detecção de bordas em imagens mais complexas.

- *Sobel edge detection* [64], semelhante ao operador *Prewitt*, mas utilizando uma máscara convolucional com pesos ajustados para dar maior ênfase aos *pixels* centrais. Essa abordagem proporciona uma suavização adicional, tornando o método mais robusto a ruídos em comparação com *Prewitt*.

- *Canny* [Canny 1986], um método mais avançado e robusto, composto por múltiplas etapas para detecção precisa de bordas. Primeiro, a imagem é suavizada com um filtro gaussiano para reduzir ruídos. Em seguida, calcula-se o gradiente da imagem para identificar variações de intensidade. Após essa etapa, a supressão de não-máximos é aplicada para remover bordas espúrias, garantindo que apenas os contornos mais fortes sejam mantidos. Por fim, um mecanismo de histerese com dois limiares distintos é utilizado para conectar segmentos de borda, resultando em uma detecção mais refinada e contínua.

2.2.3 Segmentação Baseada em Limiarização (*Thresholding*)

A segmentação baseada em limiarização separa regiões de interesse na imagem com base em um valor de intensidade previamente definido, classificando os *pixels* em duas categorias: aqueles cuja intensidade está acima do limiar e aqueles cuja intensidade está abaixo. Essa abordagem é amplamente utilizada por sua simplicidade e eficiência computacional.

Um dos métodos mais conhecidos é:

- *Otsu* [Otsu 1979], um algoritmo que determina automaticamente o limiar ótimo ao minimizar a variância intra-classe dos *pixels* segmentados. Esse método é particularmente útil para imagens bimodais, ou seja, imagens onde há uma distinção clara entre o objeto de interesse e o fundo. Sua principal vantagem é a ausência da necessidade de definição manual do limiar, tornando-o um método eficaz e amplamente utilizado em diversas aplicações de segmentação de imagens.

2.2.4 Métodos Híbridos e Redes Neurais

Apesar dos avanços das redes neurais na segmentação de imagens, os métodos tradicionais continuam sendo amplamente utilizados devido à sua eficiência computacional e simplicidade de implementação. Em muitos casos, técnicas híbridas que combinam segmentação tradicional com aprendizado profundo são empregadas para melhorar a precisão e a robustez do modelo.

Por exemplo, redes neurais convolucionais (*Convolutional Neural Networks* - CNNs) podem ser treinadas para refinar segmentações obtidas por métodos clássicos, corrigindo erros e aprimorando os contornos dos objetos segmentados. Essa abordagem permite unir a rapidez e a simplicidade dos métodos de Processamento Digital de Imagens (PDI) com a capacidade das redes neurais de aprender padrões complexos e se adaptar a variações nos dados. Além disso, estratégias como o uso de mapas de bordas extraídos de métodos determinísticos podem auxiliar na otimização do treinamento de redes neurais, fornecendo informações estruturadas e contribuindo para uma segmentação mais precisa e robusta.

2.2.5 Revisão dos métodos clássicos apresentados

Embora os métodos clássicos apresentados possuam grande poder de segmentação em condições específicas, a natureza do problema estudado apresenta diversas características que podem comprometer sua efetividade.

Um dos principais desafios é a homogeneidade da coloração dos grãos, que afeta significativamente os métodos baseados em detecção de bordas. Como esses métodos dependem de variações de intensidade nas extremidades dos objetos para realizar a segmentação, a ausência de contrastes acentuados dificulta a identificação precisa dos contornos.

Os métodos baseados em *thresholding* também enfrentam limitações devido ao grande número de objetos sobrepostos na imagem. A determinação de um limiar adequado para separar os objetos do fundo torna-se uma tarefa complexa, uma vez que um único valor pode não ser suficiente para segmentar todas as regiões corretamente, resultando em falhas na segmentação ou excesso de fragmentação.

Por outro lado, os métodos baseados em região, como *watershed* e *K-means*, são altamente sensíveis à presença de objetos com características semelhantes. Essa sensibilidade pode levar a resultados inconsistentes entre diferentes execuções do método, variando desde a perda de detalhes importantes na segmentação até a ocorrência de supersegmentação, onde um mesmo objeto pode ser indevidamente dividido em múltiplas regiões [Wu e Li 2022].

Dessa forma, a escolha do método de segmentação adequado deve levar em consideração as características específicas do conjunto de dados, e, muitas vezes, a combinação de diferentes abordagens pode ser necessária para obter melhores resultados.

2.3 Segmentação com redes profundas

Dadas as limitações dos métodos determinísticos de segmentação para a aplicação em estudo, investiga-se o uso de redes neurais profundas como alternativa para essa tarefa [Abdulateef e Salman 2021]. Nesse contexto, arquiteturas baseadas em Redes Neurais Convolucionais (CNNs) têm sido amplamente empregadas, com destaque para a U-Net [Traversi et al. 2021], uma rede projetada para a reconstrução de imagens segmentadas. A imagem de saída reconstruída representa os objetos segmentados, facilitando a extração de características relevantes [Huang et al. 2019].

Dentre as categorias de redes CNN frequentemente utilizadas como base para a U-Net, destacam-se:

- **VGG** [Simonyan e Zisserman 2014]: Rede com múltiplas camadas convolucionais empilhadas, caracterizada por sua simplicidade estrutural e eficiência na extração de características.

- **DenseNet** [Huang et al. 2017]: Utiliza conexões densas entre as camadas convolucionais, promovendo um melhor fluxo de gradiente e reutilização de características, o que melhora o aprendizado da rede.

- **ResNet** [He et al. 2016]: Implementa conexões residuais para evitar o problema do desvanecimento do gradiente, permitindo a construção de redes muito profundas sem perda significativa de desempenho.

Essas arquiteturas, originalmente projetadas para classificação de imagens, têm se mostrado alternativas eficientes quando adaptadas à estrutura da U-Net, contribuindo para melhorar a precisão e robustez da segmentação.

2.3.1 VGG

As redes do Grupo de Geometria Visual (VGG, do ingles *Visual Geometry Group*) propõem uma melhoria nas Redes Neurais Convolucionais (CNNs) por meio da utilização de janelas de convolução menores, reduzindo o número de parâmetros a serem ajustados durante o treinamento [Simonyan e Zisserman 2014]. Essa abordagem permitiu o aumento do número de camadas profundas sem que o custo computacional se tornasse proibitivo.

As seis principais configurações da rede VGG podem ser observadas na Figura 5.

Neste trabalho, foram utilizadas as variações da VGG com 16 e 19 camadas, que apresentam um melhor equilíbrio entre profundidade da rede e viabilidade computacional para a segmentação de imagens.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figura 5: Configuração das camadas da rede VGG, variando de 11 até 19 camadas, conforme proposto em sua publicação original.

2.3.2 ResNet

O problema do desvanecimento do gradiente já era conhecido desde as primeiras redes neurais profundas, mas tornou-se ainda mais evidente com o aumento da profundidade das arquiteturas convolucionais modernas. Com a introdução das redes VGG, que reduziram o tamanho dos filtros de convolução e aumentaram o número de camadas, a dificuldade no treinamento de redes profundas foi acentuada [He et al. 2016]. Isso ocorre porque, à medida que o gradiente é propagado de volta por muitas camadas, ele pode se tornar excessivamente pequeno (*vanishing gradient*) ou crescer descontroladamente (*exploding gradient*), dificultando a atualização eficaz dos pesos da rede.

Para mitigar esses problemas, a *Residual Neural Network* (ResNet) foi proposta, introduzindo conexões residuais, conhecidas como *skip connections*, entre camadas convolucionais [He et al. 2016]. Esse mecanismo permite que a informação seja propagada diretamente entre camadas mais profundas, reduzindo o efeito do desvanecimento e facilitando o treinamento de redes muito profundas. Além disso, as *skip connections* possibi-

litam que a rede aprenda representações residuais, tornando o treinamento mais eficiente e melhorando a capacidade de generalização.

A Figura 6 ilustra a arquitetura geral da ResNet e o uso das conexões residuais.

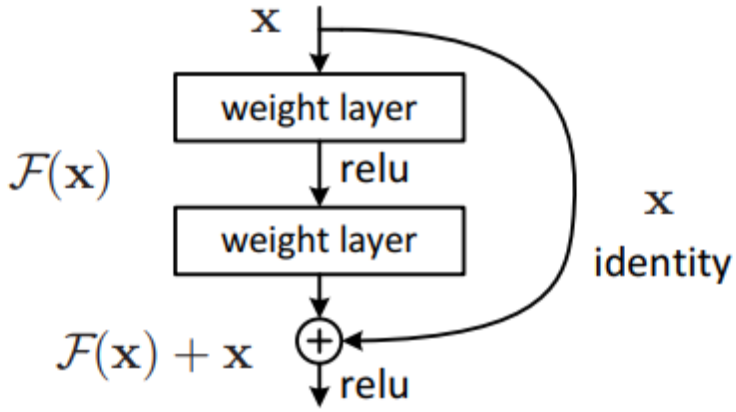


Figura 6: Arquitetura da ResNet com conexões residuais (*skip connections*).

Graças a essa abordagem, a ResNet pode alcançar profundidades significativamente maiores do que arquiteturas anteriores, como a VGG. Enquanto as redes VGG possuem algumas dezenas de camadas, a ResNet pode ser escalada para centenas de camadas sem sofrer degradação no desempenho. Isso ocorre porque as conexões residuais facilitam a propagação do gradiente durante o treinamento, permitindo que camadas mais profundas aprendam de maneira mais eficiente.

A estrutura fundamental da ResNet é baseada em blocos residuais, onde a entrada original da camada (X) é preservada e somada ao resultado da convolução subsequente. A Figura 7 ilustra um bloco de aprendizado residual, conforme apresentado no trabalho original, onde a conexão $\mathbf{X}$ *identity* representa uma *skip connection*.

A introdução da ResNet representou um avanço importante no aprendizado profundo, tornando viável a utilização de redes muito profundas em diversas aplicações, como segmentação de imagens, detecção de objetos e reconhecimento de padrões complexos.

2.3.3 DenseNet

A rede *Densely Connected Convolutional Network* (DenseNet) [Huang et al. 2017] foi desenvolvida como uma abordagem para mitigar problemas associados ao treinamento de redes neurais convolucionais muito profundas. Um dos desafios enfrentados é o fenômeno conhecido como *wash-out* [Huang et al. 2017], onde a informação pode se perder ou se tornar irrelevante após passar por um grande número de camadas entre a entrada e a saída da rede.

Para solucionar esse problema, a DenseNet introduziu um mecanismo de conexões densas (*dense connections*), onde cada camada convolucional recebe não apenas a saída

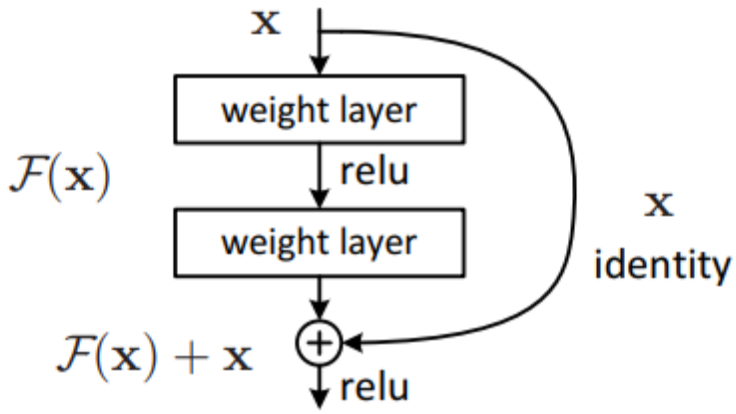


Figura 7: Bloco de aprendizado residual do trabalho original onde a conexão X *identity* representa uma *skip connection* [He et al. 2016].

da camada anterior, mas também as saídas de todas as camadas precedentes dentro do mesmo bloco. Dessa forma, a rede mantém a reutilização eficiente das características extraídas, reduzindo a perda de informação ao longo das camadas e reforçando a transmissão de atributos essenciais.

A Figura 8 ilustra a estrutura de um bloco denso com cinco camadas, conforme apresentado no trabalho original [Huang et al. 2017], demonstrando que cada camada recebe como entrada os mapas de características das camadas anteriores.

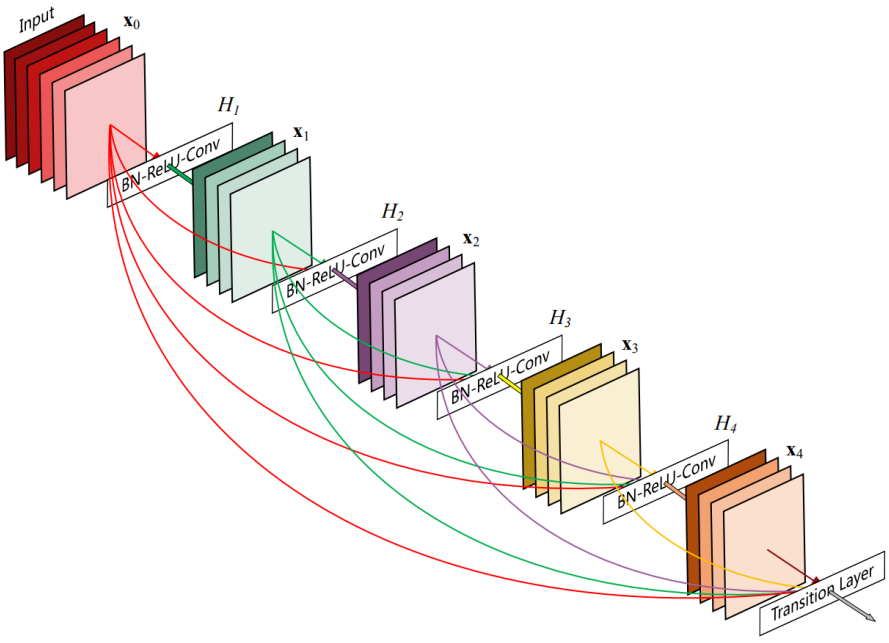


Figura 8: Bloco denso de 5 camadas, como apresentado no trabalho original [Huang et al. 2017], demonstrando que cada camada recebe todos os mapas de características anteriores como entrada.

Essa abordagem melhora a propagação da informação e o fluxo de gradiente durante o treinamento, permitindo um aprendizado mais eficiente e reduzindo a necessidade de reaprender características já extraídas em camadas anteriores. Além disso, a *DenseNet* favorece um uso mais eficiente dos parâmetros, pois promove a reutilização de características ao longo da rede.

Na arquitetura *DenseNet-UNet*, utilizada para segmentação, algumas camadas específicas, como as camadas 0, 4, 9 e 54, foram selecionadas como conexões do tipo *skip* e *bridge*, permitindo uma melhor reconstrução das imagens segmentadas.

2.3.4 U-Net

A U-Net foi originalmente desenvolvida para a segmentação de imagens biomédicas [Traversi et al. 2021]. Em seu núcleo, a U-Net é uma rede neural convolucional projetada para capturar o contexto da imagem e realizar segmentações precisas, conforme ilustrado na Figura 9.

A arquitetura da U-Net é composta por duas partes principais:

- ***Encoder***: responsável por processar a imagem de entrada, reduzindo-a a uma representação de menor dimensão contendo suas principais características.

- ***Decoder***: recebe as representações parciais do *encoder* e as expande progressivamente até reconstruir uma imagem segmentada com o mesmo tamanho da entrada.

Essas duas partes são conectadas por três *skip connections* e uma ligação direta entre o final do *encoder* e o início do *decoder*, conhecida como *bridge*. Essa estrutura permite à rede aprender padrões complexos a partir de um número limitado de imagens rotuladas.

Em geral, a U-Net recebe duas imagens como entrada: uma imagem original e sua máscara de segmentação correspondente (*ground truth*). No codificador, a imagem original passa por uma série de camadas convolucionais, reduzindo progressivamente sua resolução por meio de operações de *max-pooling*. O tamanho da imagem é reduzido do valor original (por exemplo, 512 pixels) para 64 *pixels*, permitindo a extração de características mais abstratas. Antes de cada operação de *max-pooling*, três conexões entre o *encoder* e o *decoder* enviam imagens parciais, garantindo a preservação de informações espaciais essenciais para a reconstrução da imagem segmentada.

No decodificador, o processo é invertido. Operações de *up-sampling* são utilizadas para reconstruir a imagem segmentada, unindo as representações de menor dimensão e combinando-as com as conexões *skip connections*. Essas conexões auxiliam na preservação de detalhes, evitando perdas excessivas de informação que poderiam ocorrer ao longo do processamento. O esquema da U-Net é apresentado na Figura 9.

Para melhorar a extração de características no *encoder*, redes convolucionais pré-treinadas foram adotadas. Neste trabalho, foram utilizadas as arquiteturas *VGG16,*

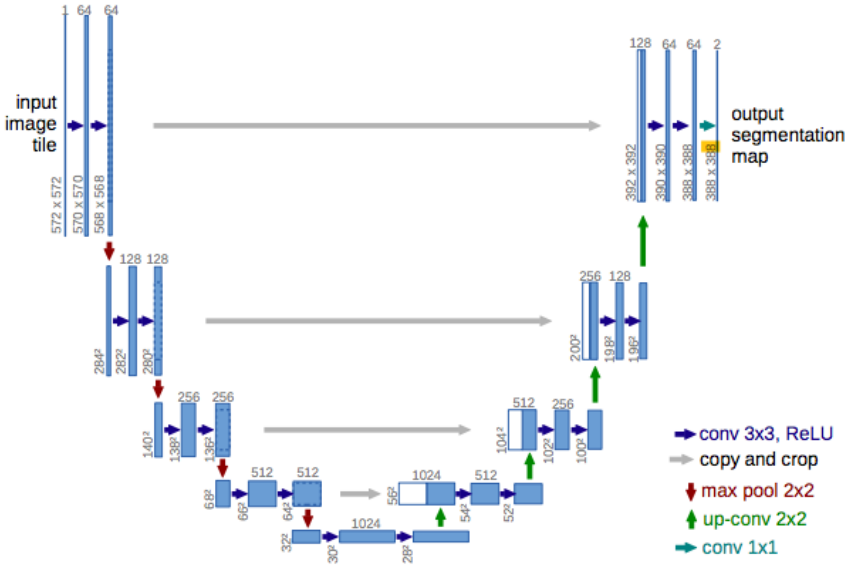


Figura 9: Arquitetura da U-Net.

VGG19, ResNet50 e DenseNet como substitutas do *encoder* original da U-Net. Essas redes foram escolhidas devido à sua popularidade, eficiência em classificação de imagens e disponibilidade de pesos pré-treinados no conjunto de dados *ImageNet*, o que acelera o treinamento e melhora a generalização do modelo.

2.4 Métricas de Avaliação

Considerando a morfologia altamente irregular dos grãos de fertilizante, que não seguem padrões geométricos consistentes e apresentam contornos complexos, abordagens baseadas na identificação de objetos por redes neurais não demonstraram desempenho satisfatório. Tais métodos, embora eficazes em contextos com objetos de formas regulares e bem definidas, enfrentam limitações significativas ao lidar com estruturas cujas fronteiras são difusas e variáveis. Em virtude dessas características, a abordagem adotada para a segmentação privilegiou a detecção das bordas dos grãos, visando capturar com maior precisão suas delimitações físicas e preservar a fidelidade morfológica das amostras.

Devido à natureza da anotação do *ground truth*, onde uma grande área branca representa os grãos e as bordas são delimitadas por linhas pretas, a avaliação da segmentação exige estratégias específicas para garantir uma análise precisa dos resultados. Para isso, os métodos de avaliação foram estruturados em duas etapas complementares.

Na primeira etapa, métricas clássicas de similaridade foram aplicadas com o objetivo de comparar os resultados das segmentações geradas pelos diferentes modelos em relação às máscaras de referência. Essa análise inicial permitiu identificar quais modelos apresentavam o melhor desempenho na preservação das características estruturais das segmentações, considerando a fidelidade da reconstrução das bordas e a manutenção da

continuidade dos contornos dos grãos.

Na segunda etapa, a análise foi voltada para a identificação individual dos grãos segmentados, permitindo a extração de informações detalhadas sobre cada um. Esse processo possibilitou registrar o tamanho de cada grão e a contagem total de grãos presentes na imagem, além de calcular o tamanho médio dos grãos. Essas métricas quantitativas fornecem uma avaliação mais aprofundada da segmentação, complementando as análises baseadas em medidas de similaridade.

Como um dos objetivos secundários deste trabalho é a análise e classificação do tamanho dos grãos, a abordagem adotada nesta etapa foi estruturada para garantir a extração precisa dessas informações. Dessa forma, a metodologia aplicada não apenas avalia a segmentação, mas também viabiliza a caracterização quantitativa dos grãos, essencial para futuras aplicações de classificação e estudo da distribuição de tamanhos.

A combinação dessas duas abordagens oferece uma análise mais abrangente da segmentação, garantindo não apenas a comparação global das imagens, mas também uma avaliação granular que considera a integridade e a individualidade dos grãos segmentados. Essa estratégia permite avaliar com precisão o desempenho do processo de segmentação, assegurando que os contornos dos grãos sejam representados de forma fiel. Isso é especialmente relevante em aplicações práticas que dependem da integridade morfológica das amostras.

2.4.1 Medidas de Similaridade

A avaliação quantitativa dos modelos foi realizada utilizando as seguintes métricas amplamente adotadas na literatura para análise de desempenho em tarefas de segmentação de imagens. A seguir, são detalhadas cada uma dessas métricas:

2.4.1.1 RMSE (Root Mean Square Error)

RMSE mede a diferença média entre os valores dos pixels das imagens segmentadas e das máscaras de referência, quantificando a mudança introduzida pelo processamento. Valores mais baixos de RMSE indicam maior similaridade entre as imagens.

2.4.1.2 PSNR (Peak Signal-to-Noise Ratio)

PSNR avalia a relação entre a potência máxima do sinal e o ruído introduzido pela segmentação. Expresso em decibéis (dB), valores mais altos (tipicamente entre 30 e 50 dB) indicam maior fidelidade na segmentação.

2.4.1.3 SSIM (Structural Similarity Index Measure)

SSIM é uma métrica que avalia a similaridade perceptual entre duas imagens, levando em consideração variações de luminância, contraste e estrutura. Seu valor varia de -1 a 1, onde 1 indica uma correspondência estrutural perfeita.

2.4.1.4 FSIM (Feature Similarity Indexing Method)

FSIM avalia a similaridade estrutural e de características entre a imagem segmentada e o *ground truth*, baseando-se na congruência de fase e na magnitude do gradiente. O valor do FSIM varia entre 0 e 1, sendo 1 a similaridade perfeita.

2.4.1.5 Dice Score:

Dice mede a sobreposição entre a segmentação gerada e a máscara de referência. Seu valor varia de 0 a 1, onde 1 representa uma sobreposição perfeita. Essa métrica é especialmente relevante para avaliar a precisão na identificação das bordas dos grãos, assegurando que a segmentação preserve a estrutura original.

2.4.1.6 Índice de Jaccard (IoU)

O índice de Jaccard, conhecido também como Intersection over Union (IoU, do inglês *Intersection over Union*), mede a sobreposição entre a segmentação prevista (S_p) e a segmentação real ou esperada (S_r). O IoU é calculado pela seguinte fórmula:

$$IoU = \frac{|S_r \cap S_p|}{|S_r \cup S_p|} \tag{5}$$

Quanto mais próximo de 1 for o IoU, melhor será a segmentação obtida pelo modelo.

2.4.1.7 F-measure

A métrica F-measure avalia simultaneamente precisão e revocação (*recall*) da segmentação, sendo uma média harmônica dessas duas métricas. A F-measure é dada por:

$$F\text{-}measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

onde $Precision$ representa a taxa de pixels corretamente classificados sobre todos os pixels classificados como positivos, e $Recall$ é a taxa de pixels corretamente classificados sobre todos os pixels realmente positivos.

Valores próximos a 1 indicam desempenho balanceado, com equilíbrio entre precisão e sensibilidade.

2.4.1.8 Distância de Hausdorff

A distância de Hausdorff avalia a similaridade dos contornos entre duas imagens segmentadas, representando a maior distância mínima entre os pontos pertencentes aos dois contornos (segmentação real e predita). Ela é definida como:

$$H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b) \right\} \tag{7}$$

em que A e B são os conjuntos de pontos dos contornos real e predito, respectivamente, e $d(a, b)$ é uma métrica de distância Euclidiana entre os pontos a e b. Valores menores indicam maior precisão na segmentação.

Essas métricas permitem avaliar a qualidade das segmentações sob diferentes perspectivas, possibilitando uma análise mais robusta e abrangente dos resultados obtidos.

2.5 Data Augmentation

Dada a limitação no número de imagens disponíveis para a construção dos conjuntos de dados, diversas técnicas de *data augmentation* têm sido amplamente exploradas na literatura para aumentar a variabilidade dos dados de treinamento. No entanto, considerando a necessidade de preservar a proporção dos grãos e garantir a correspondência exata com suas respectivas máscaras, foram selecionadas três principais abordagens [Shorten e Khoshgoftaar 2019, Wen et al. 2021]:

- Espelhamento horizontal e vertical (*flipping*): As imagens e suas máscaras correspondentes são refletidas tanto horizontal quanto verticalmente, ampliando a diversidade dos dados sem alterar a estrutura dos grãos. Cada imagem original gera quatro novas variações, contribuindo para a robustez do modelo frente a diferentes orientações dos objetos segmentados.

- Rotação (*rotation*): As imagens são rotacionadas em 90 graus para a direita e para a esquerda, criando três novas variações para cada imagem original. Essa técnica melhora a capacidade da rede neural de generalizar para diferentes orientações dos grãos, tornando-a mais resistente a variações angulares durante a inferência.

- Recorte aleatório (*random crop*): Pequenas janelas de tamanho reduzido em relação à imagem original são extraídas e armazenadas como novas amostras. O número total de imagens geradas depende do tamanho da janela de recorte em relação à imagem original. Essa técnica permite aumentar significativamente a quantidade de dados disponíveis, criando variações com regiões parcialmente sobrepostas, sem comprometer a correspondência entre as imagens e suas respectivas máscaras.

Essas estratégias foram adotadas com o objetivo de ampliar a diversidade do conjunto de dados, permitindo um treinamento mais robusto do modelo sem comprometer a fidelidade da segmentação.

2.6 Trabalhos relacionados a segmentação densa usando algoritmos clássicos

A técnica PDI ACED foi aplicada para identificar a distribuição de tamanho de nanopartículas em [Meng et al. 2018], onde foi capaz de obter bons resultados usando filtro

de borda *Canny* e transformada de *Hough* circular. No entanto, para realizar a técnica de limiarização, as imagens foram totalmente dependentes do fundo presente na imagem, o que torna esse método menos viável em uma vista superior de uma esteira.

Já em [Maiti et al. 2017], a fim de analisar a análise granulométrica de minerais de agregados graúdos com tamanhos entre 63,0 mm e 9,5 mm, a massa das partículas foi estimada com base em características extraídas por técnicas de PDI. No entanto, o método aplicado não é sensível a agregados graúdos menores que 22,4 mm. Dessa forma, este método não pode ser aplicado a partículas de fertilizantes, onde a faixa estudada está entre 1mm e 4,6mm.

Um algoritmo *Otsu* melhorado foi usado para segmentar casos onde os objetos são difíceis de segmentar do fundo em [Zhan e Zhang 2019]. Ainda, em [Ma et al. 2020], foi apresentado um algoritmo de segmentação de grãos totalmente recobertos e com bordas rasas. A proposta faz uso de redes neurais para classificar minerais em uma esteira transportadora.

Em [Igathinathane et al. 2009], foi proposto o uso de computação visual para realizar a análise de distribuição de partículas, medindo o diâmetro de Ferret. Foi escolhido o uso de visão computacional para substituição do peneiramento devido a um efeito de queda, onde um granulado que possui um comprimento superior a sua largura, pode cair em uma peneira fina, quando granulado entra de ponta. Essa abordagem garante uma separação baseada em comprimento sem inconsistências observadas no peneiramento mecânico. Embora o método obtenha bons resultados, os experimentos utilizam amostras esparsas, o que inviabiliza a aplicação direta do método no caso dos fertilizantes em uma esteira.

A caracterização da morfologia dos fertilizantes utilizando análise de imagens foi verificada em [Le et al. 2018]. Neste trabalho foram definidas e implementadas as melhores condições para capturar imagens de grãos. Entre estas definições, a resolução ideal deve possuir as proporções de 34 *pixels* por mm. O sistema foi validado utilizando esferas de plástico com tamanho conhecido, e embora tenha obtido resultados aceitáveis na análise granulométrica dos fertilizantes de diversas composições, a proposta utiliza a análise de imagens de grãos esparsos, que impossibilita seu uso direto nesta pesquisa.

Em [Igathinathane et al. 2008] é proposto um método para contar partículas que retorna as dimensões dos objetos a partir de uma imagem de partículas desarticuladas, identificando seus formatos e determinando seu tamanho. A proposta define o eixo maior e menor do grão, permitindo uma análise mais fiel do seu tamanho.

Em Al-Thyabat [5] foi proposto o modelo de análise de conjuntos (*Nottsizer*) e sua validação em minerais em uma esteira. Neste trabalho são expostas limitações na segmentação e análise de minerais. Entre os pontos mais importantes mencionados está o fato que imagens superiores da esteira estão capturando apenas a camada superior da população, o que pode não representa toda a amostra propriamente. A análise é realizada com base estatística, e é necessário possuir dados representativos para aumentar a

confiabilidade dos resultados. Embora o método provenha bons resultados na análise das partículas em uma esteira, é necessária a definição média do tamanho das partículas a serem analisadas, o que torna-se uma restrição, visto que os fertilizantes possuem uma grande variedade de tamanho no ponto de análise desejado.

Um grande obstáculo na segmentação e classificação com o uso do processamento digital de imagens é enfrentado quando uma cena analisada contém inúmeros objetos que estão próximos ou em contato [Tyagi 2018]. Para esses algoritmos, é difícil identificar corretamente cada alvo a ser segmentado, pois a sobreposição dos objetos forma uma composição homogênea, devido à falta de fundo para fornecer o contraste Yang [72]. Métodos clássicos para segmentar objetos como Otsu [Otsu 1979], k-means [Lloyd 1982] e Canny [Canny 1986] dependem do contraste entre o objeto e o fundo para segmentá-lo adequadamente.

2.6.1 Metodos de segmentação estado da arte

O objetivo central deste trabalho é desenvolver um método eficaz para a segmentação de fertilizantes em imagens, tarefa que apresenta diversos desafios específicos. Entre eles estão o alto grau de sobreposição dos grãos, o grande número de objetos simultaneamente presentes nas imagens, a homogeneidade de cor e a forma irregular dos fertilizantes. Além disso, é essencial que a técnica proposta opere a uma velocidade adequada, preferencialmente próxima ao tempo real, mesmo diante de condições adversas como *motion blur* e poeira, fatores que frequentemente afetam a qualidade das imagens adquiridas.

Nesta seção serão discutidos brevemente os métodos estado da arte em segmentação aplicados ao problema descrito, destacando-se suas principais contribuições e limitações.

Método	**Sobreposição**	**Cor homogênea**	***Blur***	**Superfície coberta**	**Dados de treino**	***Fine-tuning***
SAM	5	3	2	4	Muito alto	Médio
Mask2Former	5	2	3	2	Alto	Médio
SegFormer	4	4	2	4	Médio	Baixo
OneFormer	5	3	3	4	Alto	Médio
SegDiff	3	3	4	2	Alto	Alto

Tabela 2: Comparação entre métodos *estado da arte* para segmentação

2.6.1.0.1 *Segment Anything Model* (SAM)

O *Segment Anything Model* (SAM) é uma abordagem inovadora para segmentação proposta por Kirillov et al. [Kirillov et al. 2023]. SAM foi desenvolvido com o intuito de segmentar praticamente qualquer objeto em imagens com interação humana mínima. Seu funcionamento baseia-se em um *Transformer* treinado em um vasto conjunto de dados anotados semi-automaticamente, permitindo boa generalização em diferentes contextos. SAM destaca-se por fornecer máscaras precisas mesmo em situações desafiadoras, exibindo flexibilidade e eficiência em tarefas gerais de segmentação.



Figura 10: Segmentação perfeita em objetos distintos SAM



Figura 11: Segmentação limitada em objetos similares

Entre as vantagens do SAM estão sua alta capacidade de generalização, habilidade para segmentar objetos desconhecidos e robustez em diferentes tipos de imagens. No entanto, para uma detecção customizada, o modelo necessita de um grande volume de dados para treinamento e pode enfrentar dificuldades ao segmentar objetos muito pequenos ou finos, especialmente em imagens complexas ou ruidosas. E, embora possua alta capacidade de segmentação de diferentes objetos, quando utilizado para segmentar múltiplos objetos similares, tende a ignorar alguns objetos, como pode ser visto nas figuras 10 e 11.

2.6.1.0.2 *Mask2Former*

O *Mask2Former*, proposto por Cheng et al. [Cheng et al. 2022], é um método avançado para segmentação que busca unificar segmentação semântica, de instância e panóptica em um único modelo. Baseado em uma arquitetura *Transformer*, o *Mask2Former* realiza predições simultâneas de máscaras e categorias de objetos por meio de consultas dinâmicas, gerando segmentações altamente precisas. A inovação reside em sua eficiência ao lidar com imagens complexas contendo múltiplos objetos, inclusive com sobreposições e contornos detalhados.

Os pontos fortes do Mask2Former incluem sua habilidade de integrar diferentes ti-



Figura 12: Segmentação Mask2Former



Figura 13: Segformer

pos de segmentação em um único modelo, sua alta precisão em cenários complexos e eficiência no processamento de imagens com múltiplos objetos, como pode ser visto em Figura 12. Contudo, uma limitação significativa é seu alto consumo computacional devido à complexidade do *Transformer* e à necessidade de treinamento extensivo.

2.6.1.0.3 *SegFormer*

O *SegFormer*, proposto por Xie et al. [Xie et al. 2021], é um modelo baseado em *Transformers* projetado especialmente para segmentação semântica. Destaca-se pela arquitetura simples, integrando mecanismos hierárquicos de atenção e fusão de características em múltiplas escalas. Em casos de superfícies contendo muitos grãos similares e sobrepostos, o *SegFormer* é capaz de capturar detalhes minuciosos, facilitando a distinção entre objetos adjacentes como pode ser visto na Figura 13. Entre as vantagens estão seu ótimo desempenho e eficiência computacional, consumo reduzido de recursos e boa generalização. Contudo, pode enfrentar limitações em situações de extrema sobreposição que exijam detalhamento extremamente preciso para separar os objetos corretamente.

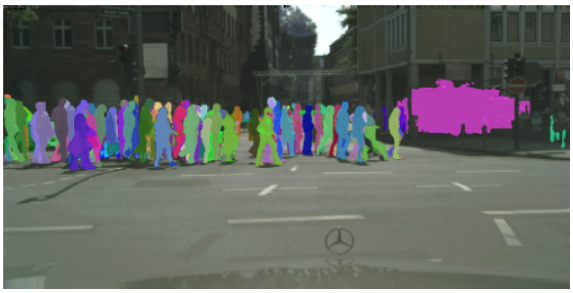


Figura 14: Oneformer

2.6.1.0.4 *OneFormer*

O *OneFormer*, desenvolvido por Jain et al. [Jain et al. 2022], é um modelo versátil que unifica segmentação semântica, de instância e panóptica em um único *framework* baseado em *Transformers*. O uso de consultas universais adaptativas permite ao *OneFormer* ajustar automaticamente seu comportamento a diferentes contextos de segmentação, o que é especialmente útil em cenários com superfícies densas com objetos sobrepostos. Essa adaptabilidade ajuda na correta distinção e classificação dos objetos em situações complexas como pode ser visto na Figura 14. Entretanto, suas principais vantagens incluem versatilidade, eficiência computacional e bom desempenho em diversos cenários com objetos distintos. Porém, limitações como alto custo computacional e a necessidade de grandes volumes de dados para treinamento são pontos negativos, e dificuldades em ambientes homogêneos acaba prejudicando a eficiência deste modelo no caso de estudo base.

2.6.1.0.5 *SegDiff*

O *SegDiff*, proposto por Amit et al. [Amit et al. 2022], utiliza modelos de difusão para realizar segmentação, gerando máscaras através de refinamento iterativo baseado em ruído gaussiano. Em casos de superfícies contendo múltiplos grãos similares e alta sobreposição, o processo iterativo do *SegDiff* pode gerar segmentações detalhadas e precisas, refinando gradualmente as fronteiras entre objetos próximos veja Figura 15. Seus pontos fortes incluem lidar bem com ruído e sobreposições complexas. No entanto, as principais desvantagens são seu elevado custo computacional e maior tempo de processamento, tornando-o menos adequado para aplicações que exijam segmentação em tempo real.

2.6.2 Considerações finais em relação aos modelos em estado da arte

A análise dos modelos de segmentação no estado da arte mostrou que cada abordagem possui vantagens e limitações específicas, dependendo do contexto da aplicação e das características visuais dos objetos a serem segmentados.

Os modelos **Mask2Former** e **OneFormer** demonstraram grande capacidade de

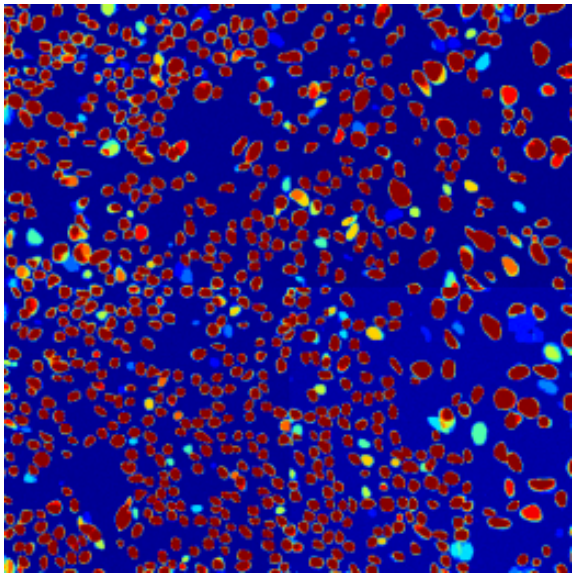


Figura 15: segdiff

segmentação, especialmente em ambientes heterogêneos, onde há variação de cores, texturas e iluminação. Isso ocorre porque esses modelos utilizam *transformers* para capturar relações contextuais entre os objetos. No entanto, quando aplicados em ambientes homogêneos, onde os objetos possuem características visuais muito similares, a segmentação pode ser menos precisa, pois a diferenciação entre instâncias pode se tornar desafiadora.

Os modelos **SegFormer** e **SegDiff** apresentaram segmentações com bordas bem definidas, sendo eficazes na delimitação dos objetos. O **SegFormer**, por ser um modelo baseado em *transformers leves*, se destaca na segmentação semântica, mas pode ter dificuldades em distinguir objetos sobrepostos, uma vez que não foi projetado para segmentação de instâncias. Já o **SegDiff**, por utilizar técnicas baseadas em *modelos de difusão*, pode oferecer contornos detalhados, mas sua eficiência pode variar conforme a complexidade da cena e o grau de sobreposição entre os objetos.

Por fim, o **Segment Anything Model (SAM)** apresentou os melhores resultados na segmentação de objetos com sobreposição irregular, destacando-se pela sua capacidade de adaptação a diferentes cenários. No entanto, seu desempenho é reduzido quando os objetos são visualmente homogêneos e apresentam sobreposição, pois o modelo depende fortemente de pistas visuais distintas para separar instâncias próximas. Esse problema pode ser mitigado com o uso de *prompts* mais específicos, mas ainda representa um desafio em cenários com alta similaridade entre os objetos. Considerando sua robustez e capacidade de segmentação em diferentes cenários, recomenda-se o uso do **SAM** para comparação com o método **GranSeg**, permitindo uma avaliação mais abrangente da eficiência de segmentação em cenários desafiadores.

Em conclusão, a escolha do modelo ideal depende fortemente do ambiente em que será aplicado. Modelos como **Mask2Former** e **OneFormer** são robustos para ambientes

diversos, mas podem enfrentar dificuldades em distinguir instâncias semelhantes. O **SegFormer** e o **SegDiff** se destacam na definição de bordas, mas não são os mais eficientes em cenários com sobreposição significativa. Já o **SAM** se sobressai na segmentação geral, mas pode falhar em contextos homogêneos. Dessa forma, a seleção do modelo deve levar em conta não apenas a precisão geral, mas também as especificidades do problema e as condições do ambiente de segmentação.

3 UMA METODOLOGIA PARA OBTENÇÃO DE *DATASETS* DE IMAGENS DENSAS COM SOBREPOSIÇÃO DE GRÃOS

Neste capítulo, as três abordagens propostas são descritas detalhadamente. Adicionalmente, a Figura 16 apresenta uma comparação entre imagens obtidas por simulação, em ambiente controlado e em ambiente de fábrica, exemplificando as diferenças visuais entre as bases de dados utilizadas nas análises conduzidas.

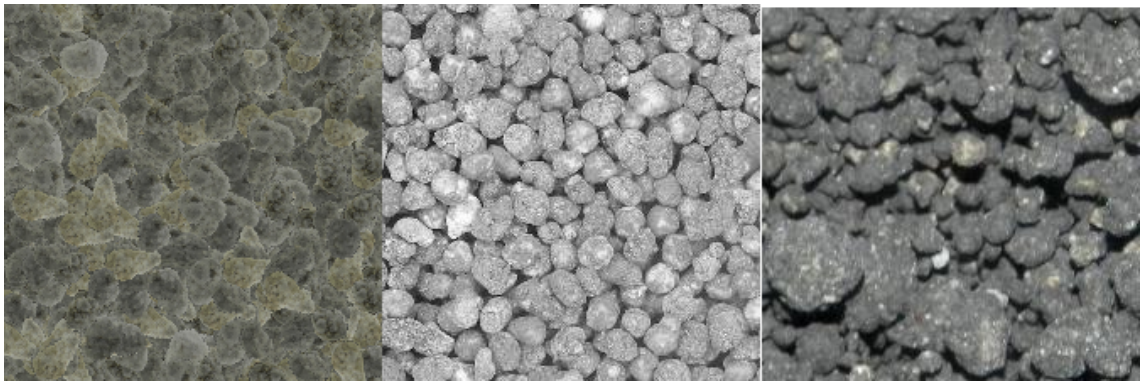


Figura 16: Comparação de imagens: Simulação, Ambiente Controlado e Fábrica.

Para conduzir o estudo, a análise e o desenvolvimento de um novo método de segmentação baseado em aprendizado de máquina, faz-se uso de dados anotados avaliar o aprendizado do modelo. No entanto, devido à inexistência de *datasets* públicos e adificuldade de obtenção de amostras adequadas e anotadas reais para esse estudo, foi desenvolvida uma metodologia para a criação de novos *datasets*. A obtenção de imagens que possibilitem a construção de *datasets* capazes de treinar e ajustar modelos de segmentação continua sendo um desafio em métodos baseados em aprendizado de máquina.

A Figura 17 apresenta uma imagem capturada em um ambiente industrial real, onde os grãos precisaram ser individualmente segmentados. Nota-se, nessa imagem, a presença de iluminação precária e heterogênea, desfoques e dificuldades na detecção de bordas, problemas que se agravam ainda mais devido à presença de poeira.



Figura 17: Imagem capturada na fábrica.

Métodos de aprendizado de máquina, especialmente aqueles baseados em *Deep Learning*, demandam para seu ajuste um grande volume de amostras e, na maioria dos casos, exigem que essas imagens sejam anotadas. Isso significa que o *ground truth*, ou seja, a máscara segmentada correspondente a cada grão, precisa ser fornecida junto com cada imagem do *dataset*.

Diante desse cenário, dois desafios principais são enfrentados. O primeiro relacionado à natureza adversa do ambiente industrial, que dificulta a obtenção de um grande volume de amostras representativas dos diferentes fenômenos de degradação e ruídos presentes. O segundo desafio refere-se à necessidade de pares anotados de *ground truth*, o que implica no uso de métodos manuais de anotação ou a adoção de técnicas automáticas, ainda em desenvolvimento.

Para lidar com essas dificuldades, duas metodologias são propostas para a geração de *datasets*. A primeira envolve o desenvolvimento de um simulador para a criação de *datasets* sintéticos, permitindo a geração de um grande número de imagens anotadas de forma automatizada e eficiente.

A segunda metodologia consiste na obtenção de *datasets* a partir de imagens capturadas em condições mais próximas à realidade denominada ambiente controlado, considerando fatores como resolução, iluminação e sombras. No entanto, esse método exige a intervenção humana para a anotação dos dados, o que limitou a quantidade de imagens geradas.

3.1 Simulador

Diante das dificuldades em obter um conjunto de dados expressivo para a comparação, análise e treinamento dos modelos, torna-se necessário o uso de imagens simuladas para representar os mais diversos cenários, incluindo casos extremos que não ocorreriam em uma fábrica operando em condições ideais. Exemplos desses casos incluem níveis de ruído acima do normal ou uma produção descalibrada, resultando em um excesso de grãos muito grossos ou muito finos. Com essa abordagem, é possível avaliar e determinar o método mais robusto a ser aplicado na fábrica.

Para isso, propõe-se o Simulador de Superfície de Grãos, cuja ideia principal é gerar um conjunto de dados simulados que represente a vista superior de um conjunto de grãos em uma esteira transportadora. Cada imagem consiste na composição sobreposta de um grande número de pequenas amostras de imagens de partículas individuais de grãos.

O primeiro passo na construção do Simulador de Superfície de Grãos foi a aquisição de imagens e dados de grupos de grãos de fertilizantes. Todos os grãos utilizados neste estudo pertencem a uma mistura do fertilizante NPK 32-05-04, composta por 5 partes de N, 32 partes de P_2O_5 e 4 partes de K_2O_5. As amostras foram coletadas diretamente da esteira transportadora após o processo de granulação, o que significa que os grãos não possuíam coloração adicional proveniente de revestimentos de óleo.

Para garantir a fidelidade da simulação, foi utilizada uma quantidade representativa de grãos de diferentes tamanhos (neste estudo, 40 grãos foram analisados). A massa de cada grão foi medida com uma balança de alta precisão, enquanto seus diâmetros foram determinados por meio de um paquímetro eletrônico. Além disso, imagens individuais de cada grão foram capturadas com uma câmera Nikon P500, com resolução de 4000×3000 *pixels*.

Com os dados coletados, o algoritmo inicia a etapa de pré-processamento. O fundo de cada imagem é removido, e o objeto em primeiro plano é centralizado. Em seguida, o grão é ajustado para garantir que sua escala corresponda às medidas reais obtidas pelo paquímetro eletrônico. A Figura 18 apresenta exemplos de grãos pré-processados.



Figura 18: Imagem de grãos pré-processados.

Propõe-se a utilização do software de processamento de imagens *ImageJ* [Ferreira e Rasband 2012] para aferir as medidas de *Feret's* máximo e mínimo, expres-

sas em *pixels*. A medida real deve ser obtida utilizando um *paquímetro eletrônico*. Após essa etapa, as imagens são redimensionadas até que a razão entre os diâmetros de *Feret* (a média da distância entre duas linhas paralelas tangentes à projeção da partícula) se iguale à razão entre a medida real e a obtida com o *paquímetro eletrônico*.

Os dados de cada grão são classificados com base na medida do diâmetro menor real, que corresponde ao tamanho da abertura da peneira pela qual o grão passa no método de peneiramento. Para a categorização dos grãos, são utilizadas oito classes baseadas no tamanho da peneira *U.S. Mesh* 1. Essa convenção segue as diretrizes do Manual de Métodos Oficiais para Fertilizantes Minerais, Orgânicos e Corretivos [Agricultura 2014], amplamente utilizado na indústria de fertilizantes.

Após a classificação, cada grão recebe uma imagem pareada, que contem apenas suas bordas realçadas em preto. Em seguida, os grãos são selecionados aleatoriamente de acordo com as proporções definidas pelo simulador e posicionados em uma imagem da matriz *A*. Para isso, um ponto aleatório (x,y) na matriz é escolhido, o grão é rotacionado em um ângulo aleatório e escalado conforme a proporção *pixel*/cm determinada previamente.

Além disso, uma segunda imagem, *B*, é gerada a partir das imagens segmentadas dos grãos, contendo exclusivamente os contornos dos grãos segmentados manualmente. A posição e a orientação desses contornos são mantidas idênticas às do grão colorido na imagem original *A*.

O simulador utiliza três matrizes com dimensões iguais às das imagens geradas (224×224), cada uma desempenhando um papel específico na organização dos grãos na imagem.

- **Matriz A (Imagem RGB):** Essa matriz armazena os valores de cor (*RGB*) dos grãos posicionados na imagem. Cada *pixel* contém a informação visual correspondente ao grão ou ao fundo.

- **Matriz B (Máscara Binária):** Essa matriz representa a segmentação dos grãos na imagem. Cada *pixel* pode assumir dois valores:

  - 0: indica a borda do grão ou a área de fundo.
  - 1: indica que o *pixel* pertence ao interior de um grão.

- **Matriz C (Mapa de IDs):** Essa matriz armazena um identificador único para cada grão na imagem, preenchendo os *pixels* da área ocupada pelo grão com seu respectivo ID. Quando um novo grão é posicionado em uma área já ocupada por outro, os valores nessa matriz são sobrescritos com o ID do grão recém-colocado, refletindo a ordem de empilhamento dos grãos na imagem simulada.

Essa estrutura permite que o simulador rastreie a localização, a segmentação e a identidade de cada grão, garantindo um controle preciso sobre a composição da imagem gerada. Além disso, mesmo que um grão seja parcialmente ocluído, sua presença ainda é registrada na matriz de IDs, enquanto grãos completamente cobertos têm seus dados substituídos.

Com a **matriz A** (*matriz de imagem*), a **matriz B** (*matriz de segmentação*) e a **matriz C** (*matriz de identificação*), o simulador contabiliza todos os grãos individuais visíveis na superfície da imagem e classifica a amostra com base na classe de maior representatividade, de acordo com a seguinte categorização:

- **Classe fina:** quando a maior parte dos grãos é capaz de passar pela peneira #8;
- **Classe ideal:** quando a maior parte dos grãos passa pela peneira #6, mas não pela peneira #8;
- **Classe grossa:** quando a maior parte dos grãos não consegue passar pela peneira #6.

As imagens geradas pelo simulador possuem resolução de 224×224 *pixels*, com cada 5 *pixels* correspondendo a um milímetro. Para cobrir toda a imagem, o simulador deposita aproximadamente 2000 grãos. No entanto, devido à sobreposição, apenas uma fração desses grãos (entre 50 e 120) pode ser observada na superfície. Essa variação depende das classes de grãos utilizadas, pois amostras compostas por grãos menores tendem a apresentar um número maior de partículas visíveis.

Na etapa final do processo, ruídos de movimentação [Zeng 2024], poeira e variações de luminosidade [Tobin et al. 2017] podem ser adicionados às cópias da *matriz de imagem*. Esse procedimento visa aproximar a imagem digital das condições reais observadas na fábrica.

Os níveis de ruído podem ser parametrizados, variando de 0,01 a 0,5. Nesse intervalo, um valor de 0,01 modificará 1% da imagem, enquanto um valor de 0,5 modificará 50% da imagem. Os resultados do uso dessas metodologias serão apresentados no próximo capítulo.

Os seguintes tipos de ruído foram considerados:

- ***Salt & Pepper*:** também conhecido como ruído de impulso, foi escolhido para simular os efeitos da poeira nas imagens simuladas, gerando valores aleatórios de *pixels* causando uma oclusão parcial. Neste filtro, é definida uma porcentagem de *pixels* a serem afetados; após a seleção desses *pixels*, seus valores são alterados aleatoriamente para 0 ou 255. O percentual de *pixels* modificados varia entre 1% e 50%, seguindo o nível de intensidade adotado neste estudo [Traversi et al. 2021].

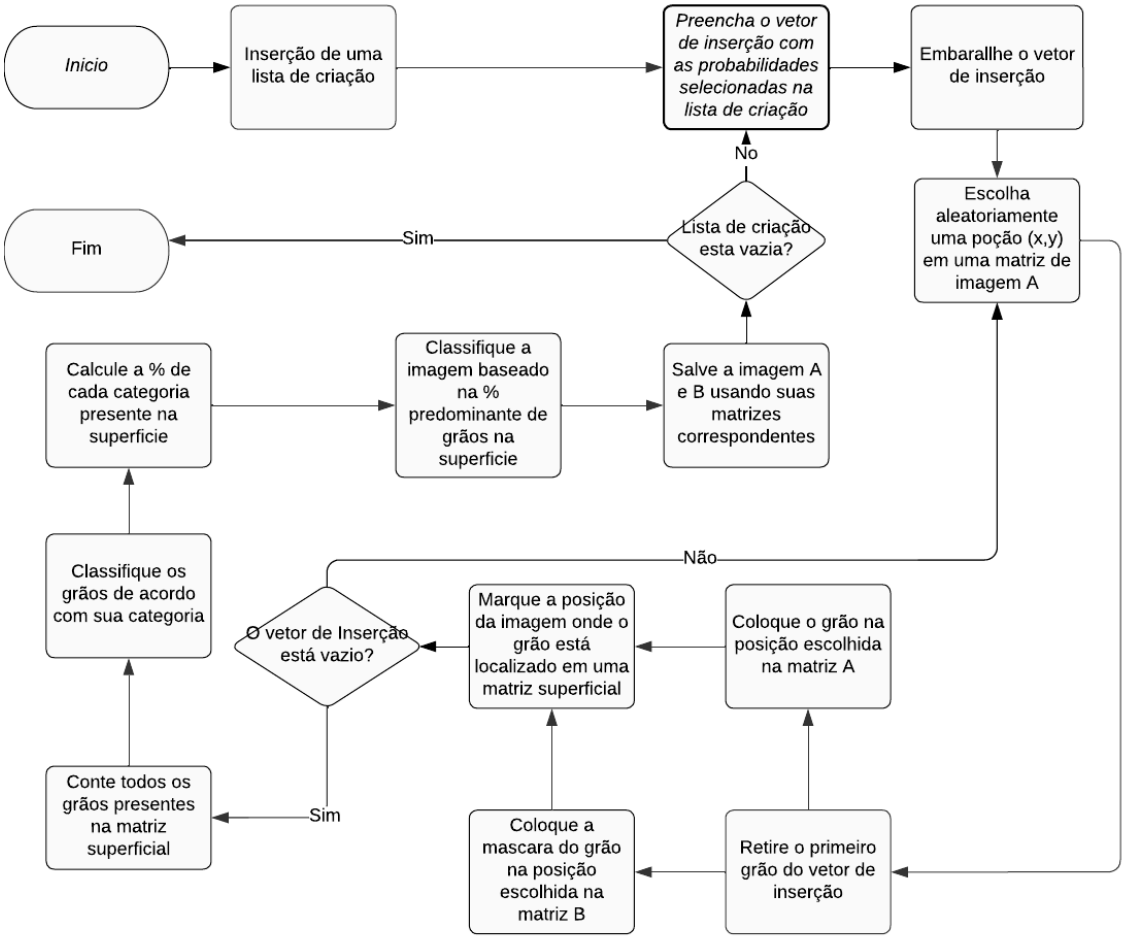


Figura 19: Fluxograma do simulador.

- ***Blur*:** o borramento de movimento e o ruído estão diretamente relacionados ao tempo de exposição durante a captura de imagens de objetos em movimento. Um tempo de exposição maior reduz o ruído, porém aumenta o *blur*; por outro lado, um tempo de exposição menor reduz o *blur*, mas aumenta o ruído da imagem. No filtro *Blur*, uma janela percorre a imagem, calculando a média dos *pixels* com seus vizinhos, o que suaviza as bordas dos objetos. O tamanho da vizinhança utilizada para a média pode variar entre 1 e 50 *pixels* [6]. Esse ruído foi escolhido para representar o efeito do movimento dos grãos sobre a esteira transportadora.

- ***Gamma*:** utilizado para ajustar os níveis de brilho na imagem. Esse filtro pode compensar variações de luminosidade, simulando imagens com superexposição ou subexposição. A equação utilizada é $I = I^{\gamma}$, onde $\gamma > 1$ escurece a imagem e $\gamma < 1$ a torna mais clara. O valor de γ pode variar de 1 a 50 na gama positiva e de 1 a -49 na gama negativa [Rahman et al. 2016]. Esse ruído foi escolhido para avaliar o impacto da iluminação insuficiente no ambiente industrial.

O fluxograma do funcionamento do simulador é apresentado na Figura 19.

3.2 *Dataset* Manualmente Anotado

A abordagem manualmente anotada segue um processo rigoroso para os *datasets* em ambiente controlado e industrial. As imagens são capturadas em formato RGB, com resolução de 4000×3000 pixels, e passam por um processo de segmentação manual detalhado, que pode demandar até 8 horas por imagem. A principal distinção entre os conjuntos de dados está na metodologia de aquisição das imagens e na complexidade da segmentação.

No *dataset* em ambiente controlado, grãos variados são dispostos sobre uma superfície circular de 15 cm de raio, iluminada uniformemente por um *ring light*. As imagens são capturadas com uma câmera posicionada a 10 cm de distância, garantindo iluminação e contraste ideais. Esse cenário assegura a visibilidade nítida das bordas dos grãos, facilitando a segmentação manual.

Por outro lado, no *dataset* em ambiente industrial, as imagens são adquiridas durante a queda dos grãos, um dos poucos pontos que permite a captura com visão próxima da superfície. No entanto, as restrições operacionais e a intensa vibração do ambiente impedem o posicionamento da câmera a menos de 20 cm da queda. Essas condições geram imagens mais realistas, porém com desafios significativos para a segmentação manual, devido à presença de ruído, variações de iluminação e sobreposição dos grãos.

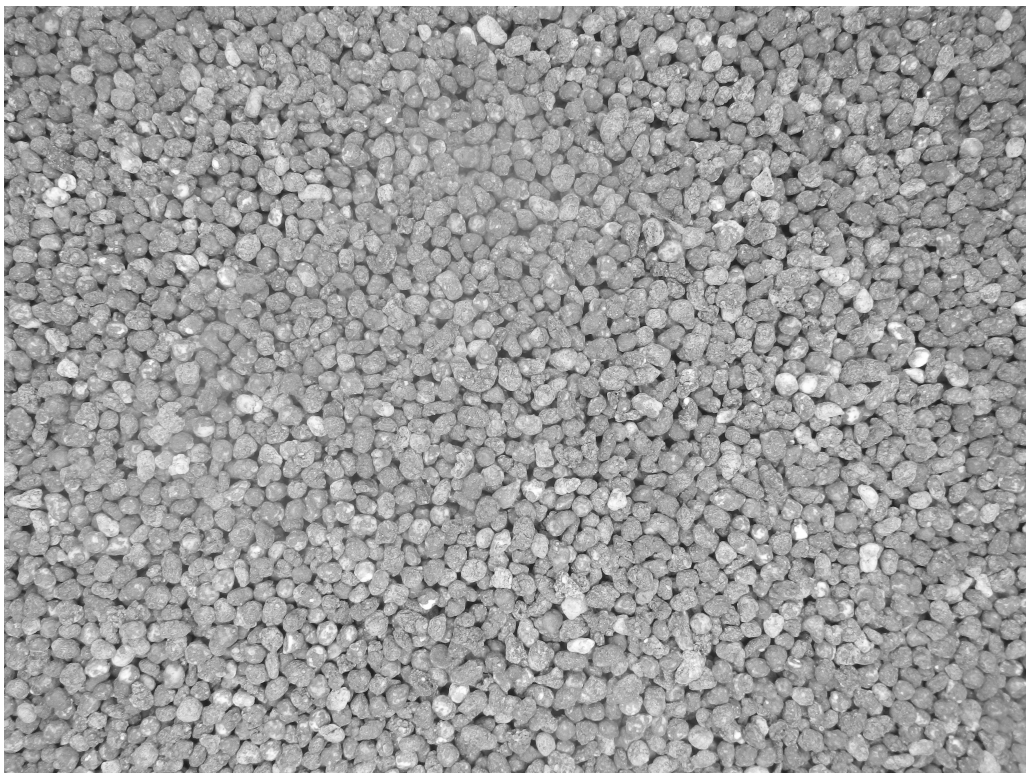


Figura 20: Imagem RGB real.

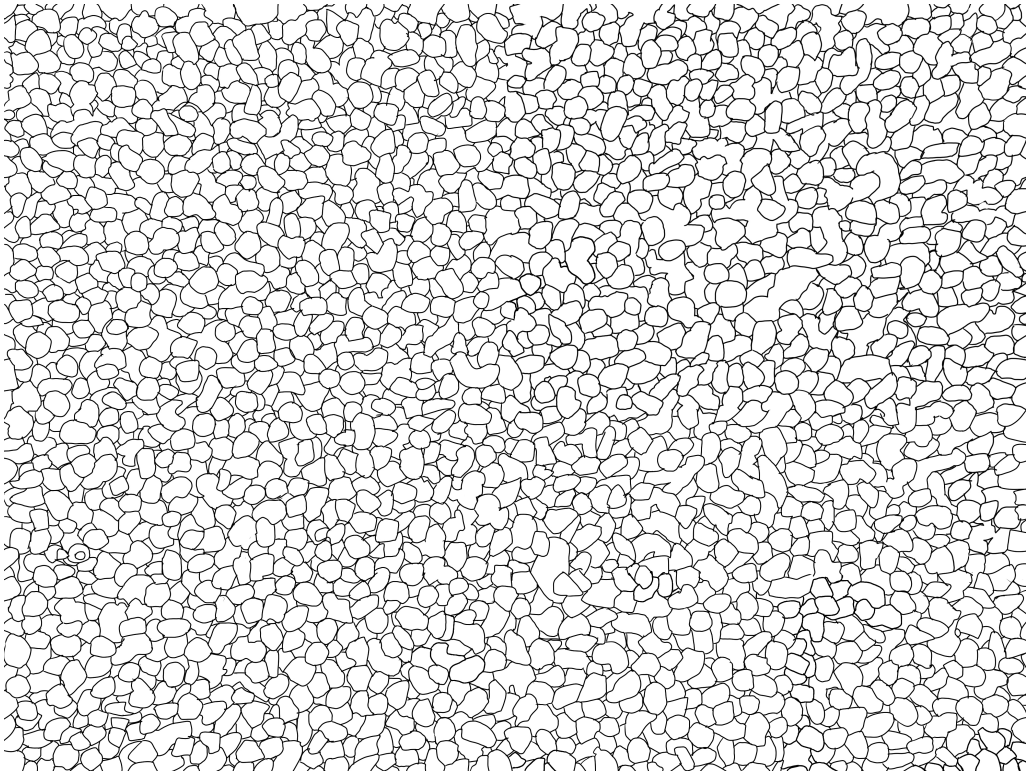


Figura 21: Máscara da imagem real.

3.3 Aplicação dos Métodos para a Obtenção dos *Datasets* para o Estudo

Nesta seção, são apresentados os conjuntos de dados (*datasets*) utilizados no estudo, com uma análise detalhada de suas características, vantagens e limitações. Inicialmente, cada *dataset* é descrito individualmente, ressaltando seus pontos fortes e fracos. Em seguida, é realizada uma comparação entre eles, considerando aspectos como brilho, contraste e distribuição de cores. Por fim, diferentes combinações de *datasets* são testadas para determinar a melhor proporção e estratégia de uso no treinamento do modelo. Essa abordagem se faz necessária devido à escassez de dados anotados no ambiente fabril.

3.3.1 Dataset Sintético

Os *datasets* sintéticos foram utilizados para avaliar a efetividade dos métodos propostos, proporcionando um grande volume de dados anotados para a realização dos experimentos.

3.3.1.1 Dataset Virtual: Análise e Teste do Simulador

O uso de um *dataset* virtual foi aplicado para a seleção de um método capaz de realizar a *segmentação* de amostras de fertilizantes sem fundo. Os *datasets* virtuais foram gerados a partir do simulador desenvolvido para este estudo.

O simulador utilizou imagens individuais de 40 grãos de fertilizantes distintos. Para a

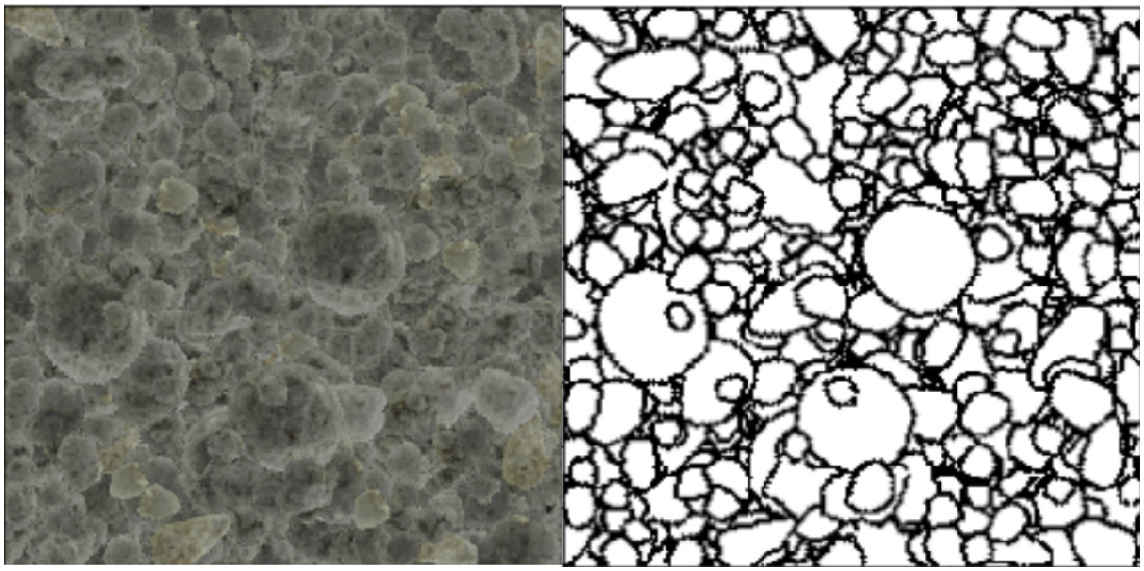


Figura 22: Figura RGB virtual e Máscara

geração de uma nova imagem, o processo inicia-se com uma superfície branca de 224 $\times$ 224 *pixels*, na qual são posicionadas de 200 a 10.000 cópias das imagens dos grãos originais. Esse procedimento resulta na criação de uma única imagem RGB e sua respectiva *máscara* apresentada na Figura 22.

Além disso, cada imagem gerada contém em seu nome as porcentagens de grãos pequenos, médios e grandes presentes na superfície. Dessa forma, é possível associar um *ground truth* detalhado da distribuição dos grãos, permitindo futuras comparações e análises quantitativas.

Tendo total controle sobre a disposição dos grãos e sua granulometria, foi possível simular diferentes cenários de produção. Entre esses, destacam-se situações com proporções semelhantes de cada categoria (30% de grãos pequenos, 40% de grãos ideais e 30% de grãos grossos), além de casos extremos em que 100% dos grãos pertencem a uma única categoria, algo incomum no contexto industrial.

Adicionalmente, com o uso de ruído *Salt & Pepper*, foi possível simular condições em que a visibilidade é reduzida em até 50%, um cenário improvável na fábrica, mas extremamente útil para avaliar a robustez do método de *segmentação*.

Por meio da aplicação de filtros de *Gamma*, foi possível identificar quais níveis de luminosidade são mais favoráveis ao processo de *segmentação*. Além disso, com o uso de *Blur*, foi possível simular a perda de informações causada pelo borramento da imagem, um problema comum quando o movimento não é adequadamente capturado por câmeras de baixa velocidade. Esse tipo de simulação permite avaliar o desempenho do método em situações onde o movimento da esteira compromete a nitidez das imagens.

Os resultados obtidos com a adição de ruídos podem ser visualizados na Figura 23.

Após a geração da imagem e de sua respectiva *máscara*, são aplicadas técnicas de rotação e espelhamento para expandir ainda mais o *dataset* criado. Esse processo per-

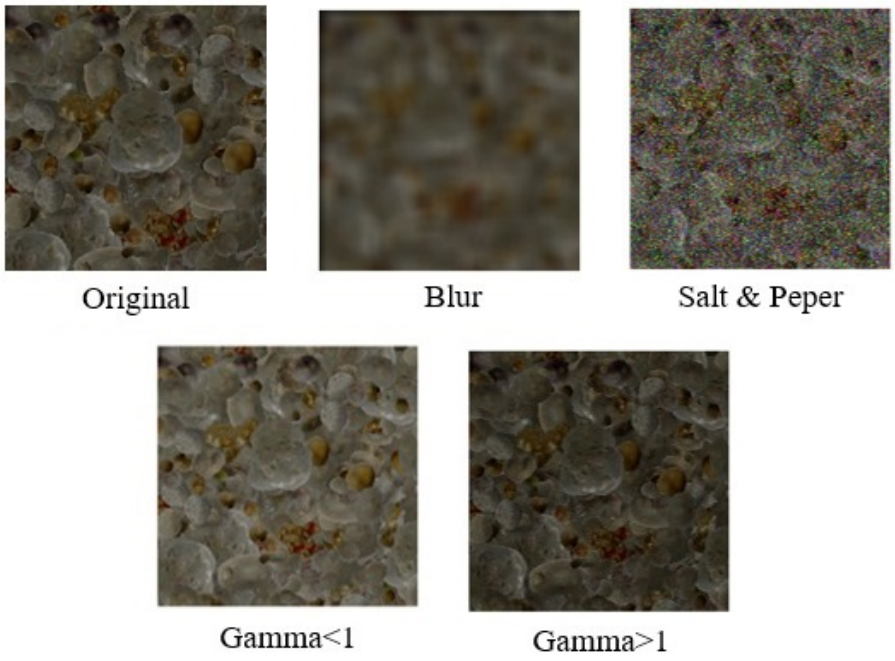


Figura 23: Ruídos de blur, salt & pepper e gamma alterado em intensidade 0.5

mite que, a cada simulação, sejam produzidas 1000 imagens e *máscaras* únicas. Com a aplicação de *data augmentation*, esse número é ampliado para um total de 6000 imagens e *máscaras*.

O ***dataset* sintético** gerado pelo simulador de superfícies oferece um alto nível de flexibilidade e escalabilidade. Sua capacidade de gerar uma quantidade virtualmente ilimitada de imagens, com variações controladas na composição dos grãos, em suas posições e nos padrões de sobreposição, torna-o adequado para a etapa de pré-treinamento de modelos de *segmentação*. Contudo, o simulador apresenta características específicas que merecem ser destacadas:

- Pontos fortes:
  - *Variabilidade infinita:* o simulador pode produzir inúmeras imagens, garantindo cobertura de diversos cenários, desde grãos espaçados até densamente empilhados;
  - *Controle sobre a composição:* controle preciso sobre as proporções dos grãos (ex.: categorias grossas, finas ou ideais) permite cenários de treinamento direcionados;
  - *Custo-efetividade:* elimina a necessidade de coleta ou anotação manual de dados, reduzindo custos operacionais.
- Desafios:
  - *Risco de sobre-ajuste:* apesar da variabilidade, o conjunto de dados virtual é

limitado pelas imagens de grãos originais usadas na simulação, levando a um possível sobre-ajuste;

- *Realismo limitado:* a natureza simulada pode não capturar completamente as complexidades de iluminação e ruído do mundo real.

3.3.1.2 Dataset com imagens reais em ambiente controlado

Para gerar o conjunto de dados reais propôs-se o uso de amostras de fertilizantes do chão de fábrica peneiradas em 3 categorias distintas: pequenos, ideais e grossos. O conjunto de dados iniciou pela captura de 10 fotos de uma superfície totalmente coberta por grãos, usando apenas partículas com o tamanho ideal em um ambiente controlado, com iluminação difusa. Essas imagens têm uma resolução de 4000x3000 *pixels*, conforme mostrado na Figura 20. Devido à iluminação controlada, os grãos aparentam possuir cor amarelada. Em seguida, a segmentação de cada grão é extraída à mão, resultando em uma segunda imagem, também com resolução de 4000x3000 *pixels*, contendo a máscara que é utilizada na *U-net* como *ground-truth*. A máscara se refere à imagem mostrada na Figura 20 e é apresentada na Figura 21.

Para expandir o conjunto de dados, propõe-se cortes aleatórios da imagem grande. Cada recorte tem o tamanho de 900x900 *pixels*, ou seja, podemos obter 12 imagens únicas de cada imagem grande e várias imagens com algumas partes sobrepostas. Cada corte também pode ser girado 3 vezes e espelhado horizontal e verticalmente, expandindo o corte original em 6 novas imagens. Assim, com as 10 imagens iniciais, consegue-se gerar 400 cortes aleatórios e cerca de 2400 imagens variadas para treinamento de redes neurais.

A aplicação utiliza imagens de 224x224 *pixels* em rgb, pareadas com a sua máscara, são apresentadas na Figura 24. Nessa representação, o grão ideal tem um raio de tamanho entre 25 a 30 *pixels*.

3.3.2 Dataset real em ambiente relevante

O conjunto de dados do ambiente industrial é composto por 1200 imagens geradas a partir das 20 melhores imagens segmentadas manualmente, capturadas diretamente no chão de fábrica. Essas imagens são obtidas em condições desafiadoras, onde fatores como iluminação irregular, reflexos e ruídos visuais impactaram a qualidade da segmentação.

Um dos principais desafios enfrentados é o posicionamento lateral da câmera, que dificultou a segmentação precisa dos grãos localizados na parte superior esquerda das imagens. Esse problema ocorre devido à inclinação da perspectiva e à presença de sombras, que reduziram a nitidez dos contornos dos grãos nessa região.

Para mitigar esse efeito e garantir um conjunto de dados mais equilibrado, as imagens são organizadas em categorias com base na área onde os recortes aleatórios foram aplicados. Dessa forma, imagens afetadas por esse problema de segmentação são agrupadas em uma categoria específica, enquanto as demais são distribuídas conforme sua quali-

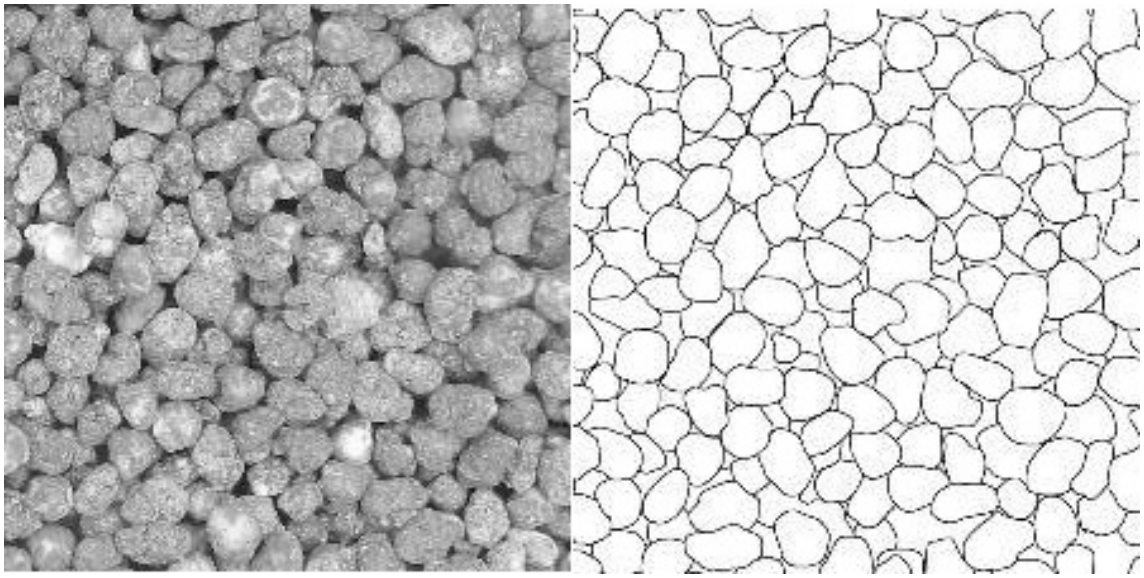


Figura 24: Recorte de 900x900 com redimensionamento para 224x224 da imagem RGB e sua mascara

dade de segmentação. Essa abordagem permite não apenas uma análise mais detalhada da robustez do método, mas também a criação de estratégias específicas para corrigir ou compensar esses desafios durante o treinamento dos modelos de *segmentação*.

A categorização estruturada do conjunto de dados assegura que ele represente uma ampla variedade de cenários industriais reais. Isso inclui desde imagens bem segmentadas, que fornecem um padrão confiável para o modelo, até aquelas que apresentam desafios complexos, simulando as condições adversas frequentemente encontradas no ambiente fabril.

A Figura 17 ilustra um exemplo de uma imagem do conjunto de dados do ambiente industrial, acompanhada de sua respectiva *máscara*, utilizada como *ground truth* no processo de treinamento e validação do modelo de segmentação.

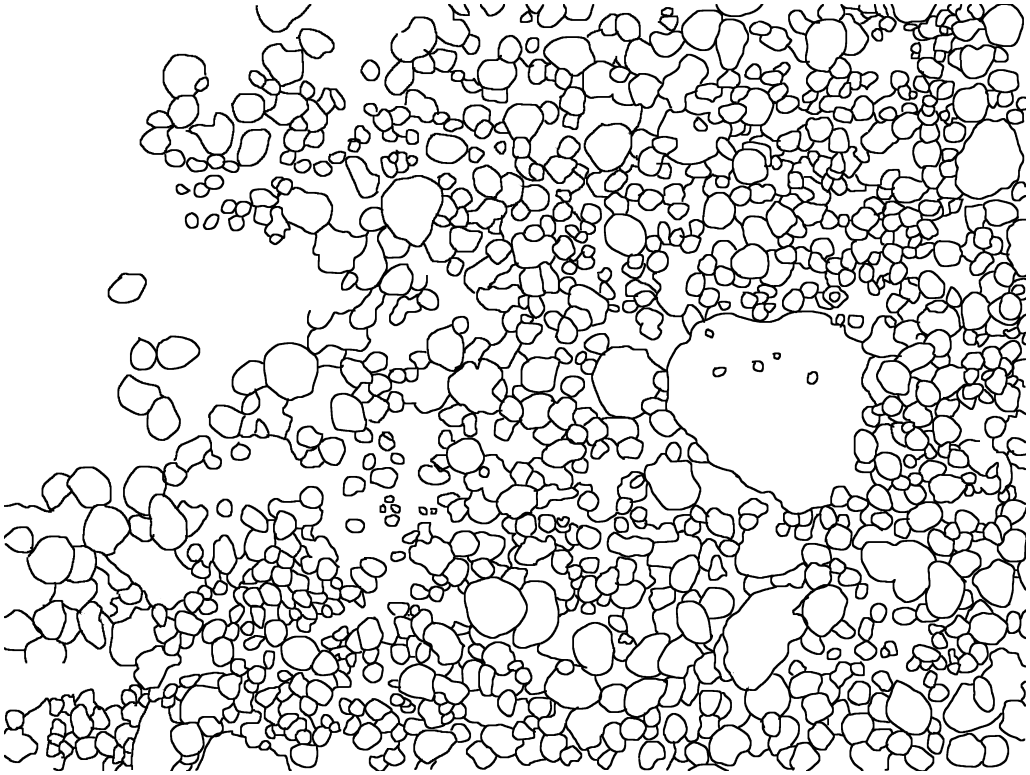


Figura 25: Mascara da imagem de fábrica.

3.4 Comparação de brilho, contraste e equilíbrio de cores

A avaliação das características visuais dos três conjuntos de dados – simulador, ambiente controlado e fábrica – foi realizada por meio da comparação de brilho, contraste e distribuição de cores, tomando como referência as imagens capturadas na fábrica. Os resultados evidenciam diferenças significativas que impactam a representatividade e aplicabilidade de cada *dataset*.

Na análise de brilho e luminosidade, as imagens da fábrica apresentaram um nível médio de brilho de 110, enquanto as geradas pelo simulador atingiram um valor próximo, de 100. Isso indica que o simulador conseguiu replicar parcialmente a iluminação do ambiente real. Em contraste, as imagens do ambiente controlado apresentaram um brilho consideravelmente maior (140), reflexo das condições otimizadas de captura, voltadas para maximizar a qualidade visual e facilitar a segmentação manual.

Na comparação de contraste, os dados da fábrica registraram um valor médio de 40, evidenciando uma variação natural de tons característica do ambiente real. As imagens do ambiente controlado apresentaram um contraste semelhante (38), sugerindo que as condições controladas mantiveram um nível de detalhamento próximo ao da fábrica. Já o simulador apresentou um contraste significativamente menor (20), indicando uma limitação na reprodução da profundidade visual e das variações tonais observadas nas imagens reais.

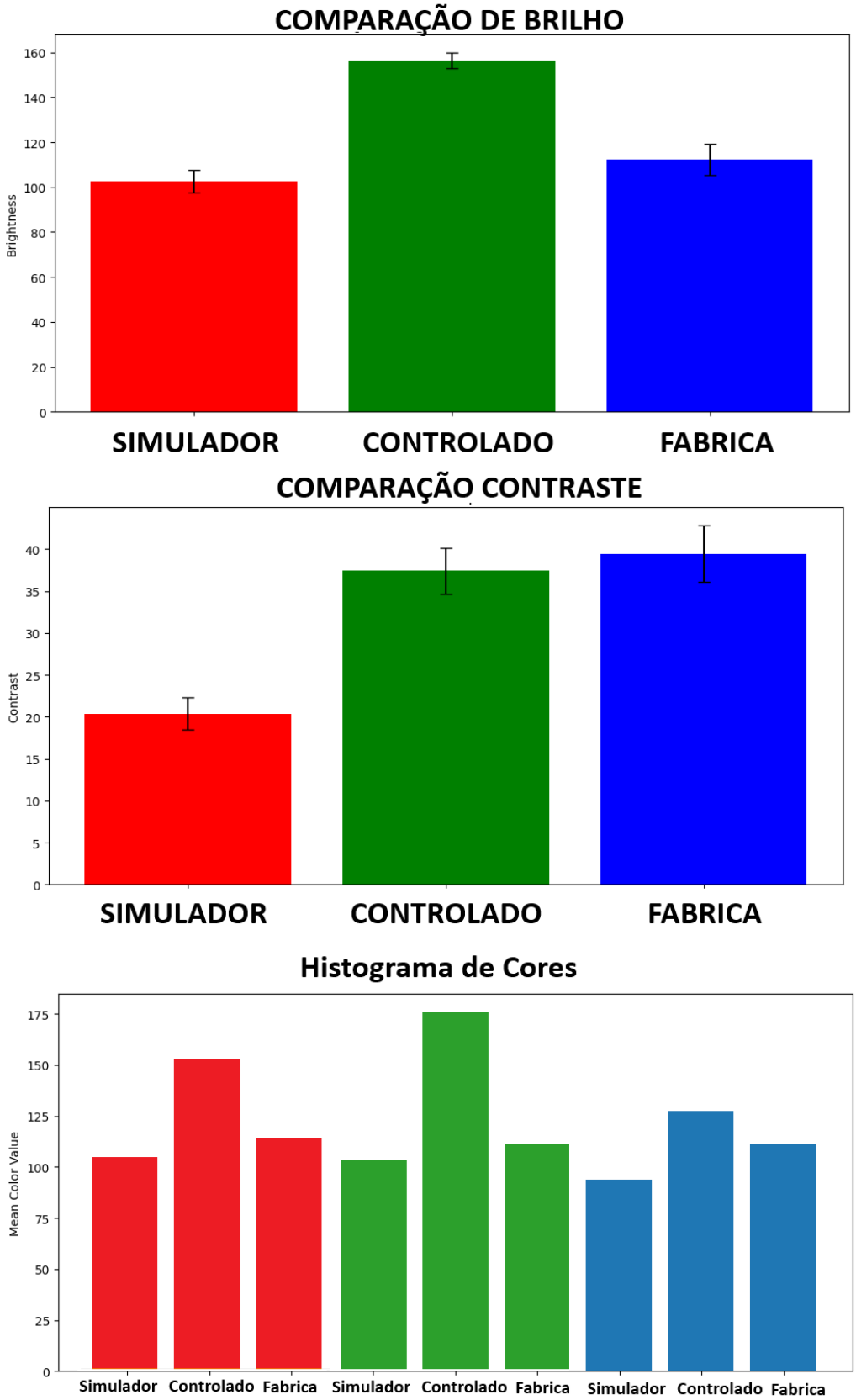


Figura 26: Comparação de brilho, contraste e histograma de cores entre os conjuntos de dados simulado, ambiente controlado e fábrica.

A análise da distribuição de cores revelou diferenças importantes entre os conjuntos de dados. Os valores médios para cada canal de cor foram:

- **Canal Vermelho**: Simulador (105), Ambiente Controlado (150), Fábrica (120);
- **Canal Verde**: Simulador (105), Ambiente Controlado (175), Fábrica (110)
- **Canal Azul**: Simulador (95), Ambiente Controlado (125), Fábrica (110); .

O ambiente controlado apresentou os maiores valores em todos os canais, refletindo as condições ideais de iluminação que realçam as cores e aumentam a separação tonal. Em contrapartida, o simulador mostrou uma menor variação de cores em relação à fábrica, enquanto o ambiente controlado se distanciou mais significativamente devido ao realce cromático provocado pela iluminação artificial.

De forma geral, o conjunto de dados da fábrica representa o cenário real, servindo como referência para avaliar a proximidade dos demais conjuntos. Enquanto o simulador conseguiu reproduzir certos aspectos, como o brilho médio, sua limitação no contraste e na definição de cores pode impactar a qualidade da segmentação. Já o ambiente controlado, apesar de apresentar uma qualidade visual superior, se distancia da fábrica em termos de iluminação e intensidade cromática, podendo influenciar a generalização dos modelos de aprendizado de máquina treinados nesse contexto.

3.5 Análise Comparativa da Eficácia de Conjuntos de Dados Sintéticos, Controlados e Reais no Treinamento de Modelos de Segmentação

Neste capítulo, é apresentada como a combinação de diferentes conjuntos de dados — sintéticos e controlados — pode aprimorar o aprendizado de modelos de segmentação, especialmente quando há disponibilidade limitada de imagens anotadas da fábrica. Idealmente, para o treinamento do modelo, o uso de imagens da fábrica devidamente segmentadas seria a escolha mais adequada, pois refletiriam com maior fidelidade as condições reais do ambiente industrial. No entanto, devido às dificuldades de acesso, captura e segmentação manual dessas imagens, foi necessário recorrer a outras abordagens para suprir essa lacuna.

Diante dessa limitação, foram utilizados dois conjuntos de dados complementares: um *dataset* sintético, gerado de forma ilimitada, que apresenta anotações ideais para a segmentação, mas possui restrições na generalização das imagens RGB, e um *dataset* obtido em ambiente controlado, que proporciona maior variação na aparência dos grãos, embora sua segmentação manual seja de qualidade inferior à do *dataset* simulado. A combinação desses conjuntos de dados busca melhorar o aprendizado do modelo, compensando as limitações individuais de cada abordagem e facilitando a segmentação

de grãos de fertilizante em condições industriais reais, onde fatores como iluminação variável, presença de poeira e o movimento dos grãos tornam a tarefa desafiadora.

Analisamos o papel desses conjuntos de dados no processo de aprendizado para tarefas de segmentação utilizando um modelo U-Net. O objetivo não é validar o desempenho do modelo U-Net em si [47], mas avaliar como diferentes combinações de conjuntos de dados influenciam a capacidade do modelo de aprender e generalizar de forma eficaz. Dado o desafio de capturar imagens do mundo real em ambientes industriais, este estudo enfatiza a importância de utilizar dados mais acessíveis, como imagens simuladas e de ambientes controlados, como alternativas viáveis às imagens capturadas diretamente na fábrica.

O modelo U-Net foi avaliado em vários cenários experimentais para determinar a importância de diferentes conjuntos de dados, tanto individualmente quanto em combinação de conjuntos de dados. Esses experimentos foram projetados para avaliar a capacidade do modelo de generalizar a partir de conjuntos de dados sintéticos e controlados para condições reais de fábrica. Os experimentos também exploraram se os conjuntos de dados sintéticos e controlados poderiam substituir os dados reais da fábrica para treinamento quando esses dados são limitados. Os conjuntos de dados utilizados nos experimentos incluíram: 3000 imagens sintéticas, 1200 imagens capturadas em condições ideais de um ambiente controlado e 400 imagens obtidas em ambiente fabril, sendo 200 imagens alocadas para ajuste fino e outras 200 usadas exclusivamente para teste.

Os experimentos foram categorizados nos seguintes ambientes:

1. **Ambiente A - Treinamento apenas com dados da fábrica:** o modelo foi treinado diretamente nas 200 imagens reais da fábrica usadas para ajuste fino. Isso serve como uma linha de base para avaliar o desempenho do modelo com dados reais limitados, sem qualquer pré-treinamento;

2. **Ambiente B - Transferência de aprendizado do ImageNet:** o modelo foi inicializado com pesos pré-treinados no ImageNet e testado diretamente nas 200 imagens reais da fábrica sem treinamento adicional. Isso testa se os recursos aprendidos de um conjunto de dados amplo podem ser transferidos para a tarefa de segmentação;

3. **Ambiente C - Generalização virtual-para-real:** treinamento com o conjunto de dados virtual (3000 imagens) e teste com 200 imagens reais da fábrica. Isso avalia a capacidade do modelo de generalizar a partir de dados sintéticos para condições reais.

4. **Ambiente D - Generalização controlada-para-real:** treinamento com o conjunto de dados do ambiente controlado (1200 imagens) e teste com 200 imagens reais da fábrica. Isso testa a eficácia dos dados controlados em replicar condições do mundo real;

5. **Ambiente E - Pré-treinamento virtual com ajuste fino:** treinamento com o conjunto de dados virtual (3000 imagens), seguido de ajuste fino nas 200 imagens reais da fábrica e teste nas 200 imagens reais da fábrica. Isso explora a utilidade do pré-treinamento com dados sintéticos para compensar a escassez de dados reais;

6. **Ambiente F: Pré-treinamento controlado com ajuste fino:** treinamento com o conjunto de dados do ambiente controlado (1200 imagens), seguido de ajuste fino nas 200 imagens reais da fábrica e teste nas 200 imagens reais da fábrica. Isso avalia se os dados controlados superam os dados sintéticos para ajuste fino;

7. **Ambiente G - Generalização mista-para-real:** treinamento com uma combinação dos conjuntos de dados virtual (3000 imagens) e controlado (1200 imagens), e teste com as 200 imagens reais da fábrica. Isso testa se a combinação dos conjuntos melhora a generalização para condições reais;

8. **Ambiente H - Pré-treinamento misto com ajuste fino:** treinamento com uma mistura dos conjuntos de dados virtual (3000 imagens) e controlado (1200 imagens), seguido de ajuste fino nas 200 imagens reais da fábrica e teste nas 200 imagens reais da fábrica. Isso avalia se a combinação dos conjuntos melhora o desempenho no ajuste fino.

Os seguintes parâmetros foram mantidos constantes em todos os ambientes:

- **Otimizador:** Adam, com taxa de aprendizado de 1×10^{-3};
- **Função de perda:** *Dice Loss*, selecionada por sua eficácia em lidar com o desequilíbrio de classes em tarefas de segmentação;
- **Métrica de avaliação:** *Intersection over Union* (IoU);
- **Tamanho do lote:** 8, equilibrando eficiência computacional e estabilidade do modelo;
- **Número máximo de épocas:** 80, com parada antecipada se a perda de validação não melhorar por 10 épocas consecutivas;
- **Aumento de dados:** Incluiu rotações, espelhamentos e recorte aleatório para melhorar a generalização do modelo.

3.5.1 Resultados da Validação

O desempenho do modelo U-Net foi avaliado em um conjunto de teste de 200 imagens reais da fábrica. Para avaliar os resultados da validação, a métrica *Dice* foi escolhida, como a segmentação das bordas envolve poucos *pixels* em comparação ao fundo, funções

de perda como *Cross-Entropy Loss* podem favorecer a predição majoritária (fundo), ignorando as bordas. A *Dice Loss* foca na similaridade entre as regiões segmentadas, equilibrando esse problema. Os resultados obtidos deste experimento são apresentados na tabela 3.

Tabela 3: Resultados experimentais.

Ambientes	Dice
A	0.72
B	0.5
C	0.77
D	0.81
E	0.96
F	0.95
G	0.83
H	0.97

A análise dos resultados experimentais apresentados na tabela 3 evidencia a influência dos diferentes conjuntos de dados e estratégias de treinamento no desempenho da segmentação. O pior desempenho foi registrado no ambiente B (*Dice* = 0,50), onde foram utilizados pesos pré-treinados do ImageNet sem treinamento adicional. Esse resultado sugere que características genéricas extraídas de imagens naturais não são suficientes para lidar com a tarefa específica de segmentação de bordas dos grãos.

No ambiente A (*Dice* = 0,72), o treinamento realizado diretamente com apenas 200 imagens reais resultou em um desempenho subótimo. Esse resultado destaca as limitações do uso de um conjunto de dados reduzido sem um pré-treinamento adequado, levando a uma baixa capacidade de generalização.

O ambiente C (*Dice* = 0,77) apresentou uma melhora em relação ao ambiente A, mas ainda assim demonstrou dificuldades na adaptação às condições reais. Esse resultado reforça que, apesar de os dados sintéticos fornecerem uma base inicial de aprendizado, o treinamento exclusivo com esse conjunto não é suficiente para capturar todas as variabilidades do ambiente real.

O treinamento com o conjunto de dados controlado (ambiente D, *Dice* = 0,81) resultou em um desempenho superior ao do conjunto virtual. Esse comportamento sugere que os dados coletados sob condições controladas possuem características mais próximas do ambiente real do que os dados puramente sintéticos.

O pré-treinamento seguido de ajuste fino mostrou-se uma estratégia eficaz. No ambiente E (*Dice* = 0,96), onde o modelo foi pré-treinado com dados sintéticos e refinado com 200 imagens reais, o desempenho melhorou significativamente. O mesmo comportamento foi observado no ambiente F (*Dice* = 0,95), onde o pré-treinamento foi realizado

com o conjunto de dados controlado antes do ajuste fino. Ambos os cenários indicam que o uso de dados alternativos para pré-treinamento é uma abordagem viável para contornar a limitação da disponibilidade de imagens reais.

A combinação dos conjuntos de dados também trouxe benefícios. No ambiente G (*Dice* = 0,83), o treinamento com um conjunto misto de dados virtuais e controlados resultou em uma melhor capacidade de generalização quando comparado ao uso isolado de qualquer um desses conjuntos.

Por fim, o melhor resultado foi obtido no ambiente H (*Dice* = 0,97), onde o treinamento inicial com dados mistos foi seguido por um ajuste fino com imagens reais. Esse resultado confirma que a integração de múltiplas fontes de dados proporciona um aprendizado mais robusto, maximizando o desempenho do modelo ao explorar características complementares de cada conjunto.

Dessa forma, os experimentos demonstram que a utilização de dados sintéticos e controlados pode reduzir a dependência de um grande volume de imagens reais, e que a combinação dessas abordagens seguida de ajuste fino com um pequeno subconjunto de imagens reais é uma estratégia altamente eficaz para melhorar a segmentação.

As seguintes inferências podem ser observadas dos conjuntos individuais:

- O conjunto de dados sintético, gerado por um simulador de superfícies, atende ao requisito de garantir controle preciso sobre variáveis como tamanho, formato e distribuição dos grãos. Essa abordagem permite extensos experimentos e ajustes finos dos modelos em um ambiente controlado e livre de ruído. No entanto, devido ao número limitado de grãos individuais, e a repetição do seu uso na montagem de novas imagens, o conjunto de dados sintético carece da variabilidade do mundo real, tornando-o menos eficaz para lidar com a imprevisibilidade dos cenários industriais;

- O conjunto de dados de ambiente controlado, capturado sob condições ideais, atende ao requisito de fornecer dados realistas, mantendo o controle sobre fatores ambientais como iluminação e ângulos da câmera. Esse conjunto atua como uma ponte entre os dados sintéticos e reais, introduzindo texturas e padrões mais realistas, minimizando o ruído. Apesar de seus benefícios, ele não replica completamente a complexidade e os desafios encontrados em ambientes industriais reais, o que pode levar a limitações na generalização do modelo;

- O conjunto de dados real da fábrica atende ao requisito de representar condições operacionais reais, incluindo ruído e variabilidade inerentes, como poeira, iluminação irregular e movimentação dinâmica dos grãos. Este conjunto de dados fornece a representação mais autêntica do ambiente de aplicação, garantindo que modelos treinados ou ajustados com ele sejam mais adequados para implantação em cenários industriais. No entanto, a coleta e anotação de dados reais da fábrica

são trabalhosas e custosas, e a variabilidade e o ruído inerentes podem exigir pré-processamento ou ajustes adicionais para alcançar o desempenho ideal.

3.5.2 Considerações finais em relação a combinação de datasets

Os experimentos realizados evidenciam que, embora conjuntos de dados sintéticos e controlados sejam valiosos para o pré-treinamento, a incorporação de dados reais é indispensável para alcançarum bom desempenho na segmentação. Modelos treinados exclusivamente com dados sintéticos ou de ambientes controlados apresentaram perdas de Dice de 0,77 e 0,81, respectivamente. Contudo, o ajuste fino com apenas 200 imagens reais da fábrica, após o pré-treinamento, elevou a perda de Dice para 0,96 e 0,95, respectivamente. Notavelmente, a combinação de conjuntos de dados sintéticos e controlados, seguida de ajuste fino com dados reais, resultou na melhor performance, com perda de Dice de 0,97. Em contraste, o treinamento apenas com 200 imagens reais resultou em uma perda de Dice de 0,72, e a transferência direta de aprendizado do ImageNet obteve uma perda de Dice de 0,50. Portanto, conclui-se que, embora dados sintéticos e controlados sejam recursos valiosos para pré-treinamento, eles não substituem completamente os dados reais. Entretanto, sua combinação com uma quantidade limitada de dados reais pode reduzir significativamente a necessidade de grandes volumes de dados reais, sendo especialmente benéfica em contextos industriais onde a coleta de dados reais é onerosa ou complexa.

4 AVALIAÇÃO COMPARATIVA DE *ENCODERS* NA U-NET EM CONJUNTO COM IMPACTO DE ENTRADAS RGB E BINÁRIAS NA SEGMENTAÇÃO DE IMAGENS

Este capítulo tem como objetivo identificar a arquitetura de rede neural convolucional (CNN) mais adequada para compor o encoder em uma rede encoder-decoder semelhante à U-Net. A partir dessa análise, será desenvolvida a rede GranSeg, que será apresentada posteriormente.

Para essa avaliação, investigam-se diferentes arquiteturas de encoder, analisando sua eficiência na segmentação das bordas dos grãos sob diferentes configurações de entrada. Em particular, são comparadas imagens RGB e binárias, a fim de compreender o impacto dessas representações no desempenho do modelo. A utilização de um conjunto de dados capturado em ambiente controlado garante condições padronizadas e segmentações de alta qualidade, permitindo uma análise mais precisa da capacidade de generalização dos diferentes encoders.

Neste contexto, são avaliadas arquiteturas amplamente utilizadas em visão computacional, incluindo VGG16, VGG19, DenseNet, ResNet-50 e ResNet-101. A análise busca determinar qual dessas redes oferece a melhor representação para a tarefa de segmentação, considerando sua capacidade de extração de características e o impacto na qualidade dos contornos segmentados.

Adicionalmente, examina-se o efeito do tipo de saída na segmentação, comparando saídas em 3 canais e saídas binárias como pode ser visto na Figura 27 . Tendo como objetivo entender como essas diferentes representações influenciam a capacidade dos modelos de representar características relevantes e a segmentação das bordas dos grãos e avaliar quais redes Convolucionais apresentam o melhor aprendizado em condições ideais.

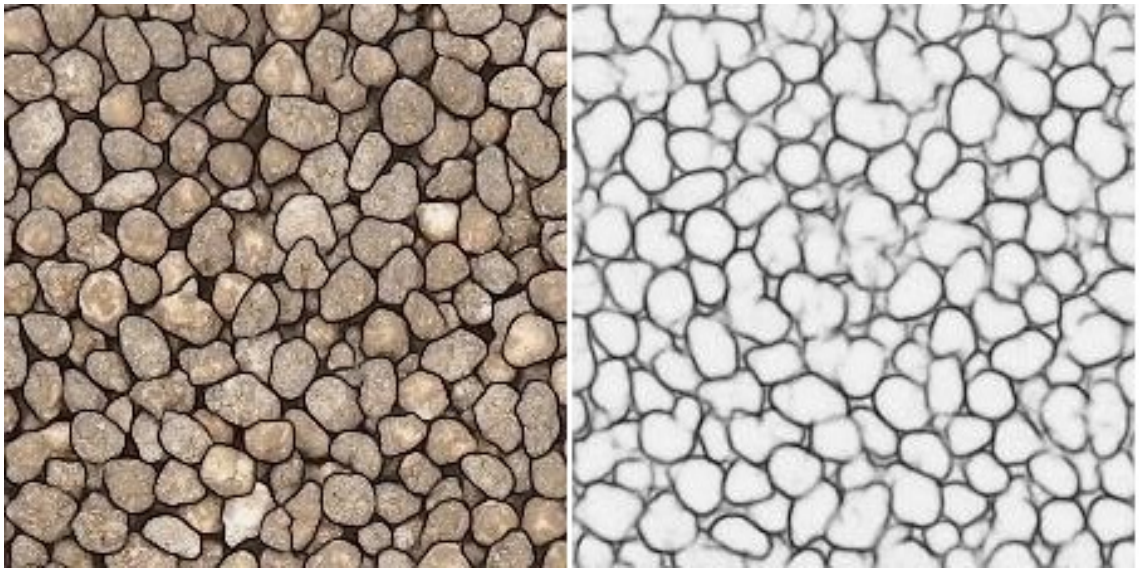


Figura 27: Output RGB da Unet e Output Binario da unet com Resnet101 encoder

4.1 Dataset

O *dataset* de ambiente controlado garante condições estáveis de iluminação e captura das imagens. O conjunto possui uma diversidade significativa de tipos e tamanhos de grãos, proporcionando desafios variados aos modelos. As segmentações anotadas manualmente fornecem alta confiabilidade nos resultados apresentados.

Um conjunto de dados com 1200 imagens, cada uma com dimensões de 224x224x3 *pixels*, foi utilizado como entrada para os experimentos. As saídas correspondentes incluíram 1200 máscaras binárias e 1200 máscaras RGB, sendo 20% das imagens aleatoriamente selecionadas como conjunto de teste e 10% adicionais separadas para validação.

4.2 Arquitetura Experimental

Os experimentos foram conduzidos considerando cinco conjuntos distintos *sets*, cada um contendo versões similares da U-Net, variando exclusivamente no *encoder* utilizado e no tipo de saída da rede (binária ou com três canais).

Os modelos avaliados receberam as seguintes nomenclaturas, refletindo os *encoders* utilizados:

- U-Net (encoder padrão);
- DenseNet-U;
- VGG16-U;
- VGG19-U;
- ResNet50-U;

- ResNet101-U.

Cada modelo foi avaliado com dois tipos de saída:

1. Segmentação binária;
2. Segmentação com três canais (RGB).

4.3 Métricas de Avaliação

As avaliações foram realizadas com métricas padrão em segmentação: erro quadrático médio (RMSE), taxa de pico sinal-ruído (PSNR), índice de similaridade estrutural (SSIM) e interseção sobre união (IoU). Essas métricas avaliam a qualidade geral e a similaridade entre os valores previstos e reais, com valores mais altos indicando melhor desempenho para PSNR e SSIM, enquanto valores mais baixos indicam melhor desempenho para RMSE.

4.4 Resultados e Discussões

Os principais resultados inferidos da Tabela 4 e 5 são apresentados a seguir:

	U-Net		DenseNet-U		Vgg16-U	
Métrica	Binário	RGB	Binário	RGB	Binário	RGB
RMSE	0.011124	0.007218	0.012793	0.009058	0.016848	0.007796
PSNR	39.331072	42.864573	38.028690	41.166782	35.551971	42.177230
SSIM	0.933506	0.961328	0.903767	0.929276	0.892137	0.954871
IoU	0.973459	0.933908	0.977080	0.922177	0.976948	0.929386

Tabela 4: Resultados U-net, Densenet-U e Vgg16-U

	Vgg19-U		Resnet50-U		Resnet101-U	
Métrica	Binário	RGB	Binário	RGB	Binário	RGB
RMSE	0.015153	0.008022	0.012084	0.008082	0.015949	0.011144
PSNR	36.410642	41.925698	38.439902	41.92756	35.963711	39.611113
SSIM	0.901323	0.952523	0.922638	0.950493	0.850051	0.892779
IoU	0.977112	0.925723	0.976938	0.934711	0.979178	0.919710

Tabela 5: Resultados Vgg19-U, Resnet50-U e Resnet101-U

4.4.1 U-Net

- **Segmentação binária**: A U-Net original apresentou uma boa precisão na tarefa de segmentação binária, com um IoU de 0,973459, demonstrando sua eficiência

em capturar detalhes complexos e delinear com exatidão os contornos dos objetos. O desempenho do modelo foi reforçado ainda por um baixo RMSE de 0,011124, indicando pequenos erros entre os valores previstos e reais. Além disso, os altos valores de PSNR (39,331072) e SSIM (0,933506) enfatizam a excelente qualidade das imagens segmentadas, garantindo uma representação fiel das regiões segmentadas;

- **Segmentação RGB**: A U-Net original manteve seu alto padrão na tarefa de segmentação RGB, atingindo um PSNR de 42,864573 e um SSIM de 0,961328. Essas métricas destacam a capacidade do modelo em preservar a fidelidade da imagem durante o processo de segmentação, sendo essa precisão corroborada pelo baixo valor de RMSE de 0,007218.

4.4.2 DenseNet-U

- **Segmentação binária**: A arquitetura DenseNet-U demonstrou um desempenho adequado na segmentação binária, igualando a precisão da U-Net original com um IoU de 0,977080. Apesar de apresentar um RMSE ligeiramente superior, de 0,012793, o modelo demonstrou capacidade em gerar máscaras binárias precisas. Os valores obtidos de PSNR (38,028690) e SSIM (0,903767) indicam uma boa qualidade de imagem, ainda que ligeiramente inferiores aos obtidos pela U-Net original;

- **Segmentação RGB**: A DenseNet-U também é uma opção interessante na segmentação RGB, obtendo um PSNR de 41,166782 e SSIM de 0,929276. Essas métricas, aliadas a um IoU de 0,922177, demonstram a eficiência do modelo em preservar a informação de cor enquanto segmenta os objetos com precisão.

4.4.3 VGG16-U

- **Segmentação binária**: A VGG16-U obteve um IoU de 0,976948 na segmentação binária, demonstrando sua eficácia em capturar padrões complexos nas imagens binárias. Além disso, apresentou um PSNR de 35,551971 e SSIM de 0,892137 que, embora ligeiramente inferiores aos outros modelos, ainda refletem uma qualidade de imagem satisfatória. A precisão geral do modelo é reforçada ainda por um RMSE relativamente baixo, de 0,016848;

- **Segmentação RGB**: A VGG16-U apresentou qualidade de imagem competitiva na segmentação RGB, com um PSNR de 42,177230 e SSIM de 0,954871. Essas métricas, juntamente com o IoU de 0,929386, destacam a capacidade do modelo em segmentar os objetos com precisão, preservando simultaneamente informações de cor e textura.

4.4.4 VGG19-U

- **Segmentação binária**: A VGG19-U apresentou alta precisão na segmentação binária, com um IoU de 0,977112, comparável à U-Net original. O modelo manteve uma boa qualidade de imagem, com PSNR de 36,410642 e SSIM de 0,901323. O RMSE ligeiramente mais elevado, de 0,015153, sugere um erro médio moderado na segmentação;

- **Segmentação RGB**: A VGG19-U apresentou qualidade de imagem competitiva na segmentação RGB, com PSNR de 41,925698 e SSIM de 0,952523. Apesar de um IoU mais baixo, de 0,925723, o modelo demonstrou capacidade em segmentar objetos com precisão, preservando os detalhes das cores.

4.4.5 ResNet50 U-Net

- **Segmentação binária**: A ResNet50-U manteve uma alta precisão na segmentação binária, alcançando um IoU de 0,976938. O modelo apresentou uma boa qualidade de imagem, com um PSNR de 38,439902 e um SSIM de 0,922638. O RMSE ligeiramente mais elevado, de 0,012084, sugere um erro médio moderado na segmentação;

- **Segmentação RGB**: A ResNet50-U demonstrou boa qualidade de imagem na segmentação RGB, obtendo PSNR de 41,92756 e SSIM de 0,950493. Essas métricas, aliadas a um IoU de 0,934711, destacaram a capacidade do modelo em segmentar os objetos com precisão, preservando simultaneamente a informação de cor.

4.4.6 ResNet101 U-Net

- **Segmentação binária**: A ResNet101-U apresentou elevada precisão na segmentação binária, obtendo um IoU de 0,979178, superando todos os demais modelos. Apesar de um RMSE mais elevado, de 0,015949, o modelo demonstrou excelente capacidade em delinear precisamente os limites dos objetos. Por outro lado, os valores mais baixos de PSNR (35,963711) e SSIM (0,850051) indicam um compromisso com a qualidade geral da imagem;

- **Segmentação RGB**: A ResNet101-U demonstrou qualidade moderada na segmentação RGB, obtendo um PSNR de 39,611113 e um SSIM de 0,892779. O IoU de 0,919710 sugere uma boa sobreposição entre as máscaras previstas e reais, embora ligeiramente inferior a outros modelos avaliados.

Os resultados mostraram que a saída RGB apresentou desempenho consistentemente superior à saída binária em termos das métricas RMSE, PSNR e SSIM em todos os conjuntos avaliados. No entanto, vale notar que a saída RGB teve uma avaliação inferior

em termos da métrica IoU quando comparada à saída binária. A métrica IoU mede especificamente a sobreposição entre as regiões previstas e reais, enfatizando a precisão na detecção de objetos ou segmentação da imagem. Essa discrepância sugere que, embora a saída RGB tenha se destacado em termos de qualidade e similaridade geral, pode ter enfrentado desafios na delimitação precisa das fronteiras ou na detecção exata dos objetos. Por outro lado, a saída binária apresentou melhor desempenho em relação ao IoU, indicando maior precisão na detecção ou segmentação dos objetos. Esses achados destacam um compromisso existente entre a qualidade geral da imagem e a precisão na delimitação dos objetos, dependendo da aplicação específica. Portanto, a escolha entre saída RGB ou binária depende do equilíbrio desejado entre a similaridade geral da imagem e a detecção precisa dos objetos.

A Figura 27 mostra os resultados obtidos pela U-Net na segmentação RGB a esquerda e o resultado obtiudo pelo modelo com a Resnet101-U. O modelo U-Net com máscara RGB apresentou o melhor desempenho nas métricas RMSE, PSNR e SSIM. Por outro lado, o modelo ResNet101-U alcançou o maior valor de IoU com máscara binária, indicando precisão superior na detecção ou segmentação dos objetos. Assim, considerando os modelos avaliados, a U-Net destacou-se como um método confiável para segmentação de grãos de fertilizantes, sendo viável para incorporação na linha de produção da indústria de fertilizantes e fornecendo uma solução para controle de qualidade sem a necessidade de interrupção na produção.

4.4.7 Considerações Finais sobre a Comparação de *Encoders* na U-Net e o Impacto das Entradas RGB e Binárias na Segmentação de Imagens

Com base nos resultados experimentais apresentados na Tabela 4, destacam-se os seguintes pontos principais:

- As saídas RGB obtiveram melhor desempenho em RMSE, PSNR e SSIM, refletindo uma melhor preservação da qualidade e fidelidade das imagens;
- Por outro lado, as saídas binárias consistentemente apresentaram valores superiores de IoU, destacando maior precisão na segmentação e delimitação dos objetos, o que é mais apropriado para aplicações que requerem melhor contorno e detecção precisa de bordas.

Esses resultados reforçam a importância de considerar a aplicação prática ao escolher entre modelos e tipos de saídas (RGB ou binária), dado o equilíbrio observado entre qualidade visual geral e precisão específica na delimitação das regiões segmentadas. Os resultados apresentados indicam claramente que o melhor *encoder* para o modelo a ser aplicado é a ResNet101-U com saída binária.

5 GRANSEG: UMA ABORDAGEM HÍBRIDA PARA SEGMENTAÇÃO DE OBJETOS SOBREPOSTOS NO CONTEXTO DA GRANULOMETRIA VISUAL DE FERTILIZANTES

A segmentação de grãos no contexto da granulometria visual de fertilizantes apresenta desafios significativos devido à alta sobreposição e homogeneidade cromática dos objetos na imagem. Para abordar essa complexidade, este trabalho propõe uma metodologia híbrida chamada GranSeg baseada na arquitetura U-Net.

Neste capítulo, será apresentada uma análise detalhada da solução proposta, abordando sua estrutura e os processos envolvidos em sua implementação. O conteúdo está organizado nas seguintes seções:

- **Visão Geral da Proposta**: Uma exploração aprofundada da arquitetura do modelo, explicando suas principais diferenças em relação à U-Net tradicional e os benefícios da combinação de múltiplos *encoders*.
- **Configuração e Treinamento da Rede**: Descrição do processo de treinamento, incluindo as configurações de hiperparâmetros, a escolha dos conjuntos de dados e as estratégias utilizadas para melhorar a convergência do modelo.
- **Pós-processamento e Extração de Características**: Apresenta as técnicas aplicadas após a segmentação para a quantificação e análise dos grãos, abordando métodos de contagem, extração de morfologia e visualização dos resultados.

Este capítulo estabelece as bases da metodologia desenvolvida, fornecendo as informações necessárias para compreender como a solução foi estruturada e implementada.

5.1 Overview da Proposta

A solução proposta neste trabalho para o desafio da segmentação de grãos no contexto da granulometria de fertilizantes, caracterizados por alta sobreposição e homogeneidade

cromática, consiste em um modelo baseado na fusão de informações advindas de métodos tradicionais de segmentação com o uso de uma na arquitetura U-Net com múltiplos *encoders* .

Diferentemente da U-Net tradicional, esta abordagem emprega duas entradas complementares: uma imagem RGB convencional, contendo informações visuais originais, e uma imagem previamente segmentada por metodos tradicionais de segmentação de imagens (por exemplos: algoritmos *prewitt [Rahmawati et al. 2021] e watershed[Sarkar et al. 2018]*), destinada especificamente a reforçar a capacidade da rede neural na identificação precisa das bordas dos objetos, como pode ser vista na Figura 28.

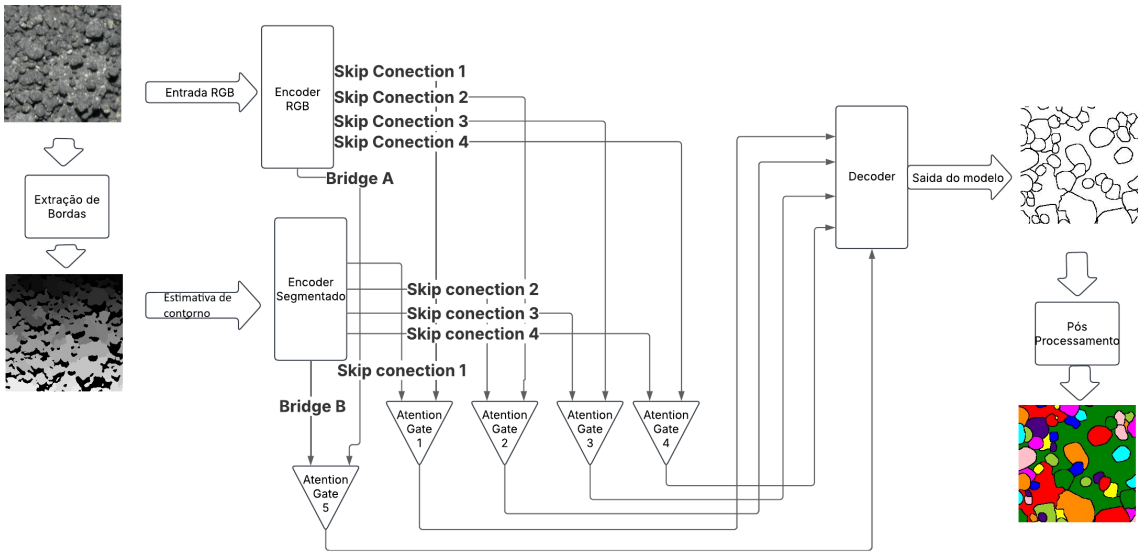


Figura 28: Modelo GranSeg

Inicialmente, a imagem original passa por um módulo de extração de bordas, gerando uma máscara responsável pela estimativa dos contornos. Em seguida, tanto a imagem original quanto a máscara extraída são processadas por seus respectivos *encoders*, que podem ser baseados em diferentes abordagens, como por exemplo, arquiteturas ResNet [He et al. 2016], VGG [Simonyan e Zisserman 2014] ou DenseNet [Huang et al. 2017]. Um *encoder* foca nas informações de cor e textura da imagem RGB, enquanto o outro enfatiza a estrutura e os limites definidos na máscara.

Os mapas de características extraídos são então combinados por meio de *skip connections* e refinados através de um *attention gate*, que destaca regiões relevantes para a segmentação. Essas informações são então encaminhadas ao *decoder*, onde ocorre a reconstrução da imagem segmentada por meio de operações de *upsampling*.

Essa abordagem integrada permite uma segmentação mais precisa, garantindo que a rede aproveite tanto as características globais quanto os detalhes finos presentes na imagem.

..

5.1.1 Principais Módulos do modelo

A seguir detalha-se cada um dos módulos que compõem o modelo.

5.1.1.1 Extração de Bordas

O primeiro módulo do modelo Granseg é responsável pela Extração de Bordas, etapa fundamental para a segmentação dos grãos na imagem. O objetivo dessa fase é realçar os contornos das partículas na imagem original em formato RGB, transformando a informação visual em uma representação mais adequada para os estágios subsequentes do processamento.

Para isso, são aplicados métodos clássicos determinísticos de detecção de bordas, como o operador de *Prewitt* [Rahmawati et al. 2021] e o *Watershed* [Sarkar et al. 2018]. O operador de *Prewitt* é um método baseado em convolução que utiliza máscaras diferenciais para estimar gradientes na imagem, realçando as transições abruptas de intensidade. Já o *Watershed* é um algoritmo de segmentação baseado na morfologia matemática, que trata a imagem como uma topografia e segmenta os objetos a partir das bacias hidrográficas identificadas.

A escolha do método de extração de bordas pode impactar diretamente a precisão da segmentação final, influenciando na qualidade da separação dos grãos e na identificação de suas fronteiras. Assim, a seleção do melhor método depende das características da imagem de entrada e das condições de iluminação.

Ao final dessa etapa, a imagem com as bordas extraídas não substitui a imagem original, mas complementa a análise ao fornecer informações estruturais cruciais. Ambas as imagens, a versão RGB original e a versão processada com as bordas realçadas, são utilizadas como entrada para as fases seguintes do modelo, garantindo uma segmentação mais robusta e precisa.

5.1.1.2 Encoder

O *Encoder* é responsável por extrair características relevantes da imagem e reduzir sua dimensão espacial. Para isso, ele é composto por três principais componentes. Primeiramente, convoluções são aplicadas para extrair características como bordas e texturas por meio de filtros. Em seguida, a normalização e a ativação são realizadas utilizando *batch normalization* e funções de ativação, como ReLU, para estabilizar o treinamento e introduzir não linearidade. Por fim, operações de *pooling*, como *maxPooling*, são empregadas para reduzir a resolução da imagem, aumentando o campo de visão e diminuindo a quantidade de parâmetros da rede.

Na arquitetura, propõe-se um encoder para tratar as imagens RGB, aproveitando as informações de cor e textura; e um segundo encoder para codificar as informações advindas das imagens segmentadas, focando em mapas binários ou multi-classe para capturar a estrutura da segmentação.

No modelo GranSeg são utilizados dois Encoders Resnet 101. Como mencionado previamente a ResNet 101 é uma rede neural convolucional profunda baseada em blocos residuais, o que permite treinar redes muito profundas sem sofrer com o problema do desaparecimento do gradiente. Sua arquitetura é compostapor um conjunto de camadas descritas a seguir:

A primeira camada da ResNet-101, realiza uma série de operações para processar a entrada. Inicialmente, aplica-se uma convolução 2D com um kernel de 7×7 e *stride* 2, reduzindo a resolução da imagem. Em seguida, a normalização em lote (*batch normalization*) é utilizada para estabilizar o treinamento ao normalizar os valores de ativação. A função de ativação ReLU é então aplicada para introduzir não linearidade ao modelo. Por fim, um *max pooling* com kernel de 3×3 e *stride* 2 é empregado para reduzir ainda mais a resolução da imagem.

A ResNet-101 é composta por diferentes conjuntos de blocos residuais organizados da seguinte forma: 3 blocos residuais com dimensões de 56×56; Res3 contém 4 blocos residuais com dimensões de 28×28; Res4 apresenta 23 blocos residuais com dimensões de 14×14; e Res5 conta com 3 blocos residuais com dimensões de 7×7.

Cada bloco residual segue uma estrutura composta por três operações convolucionais e uma conexão de atalho (*skip connection*). Primeiramente, é aplicada uma convolução 1×1 para reduzir a dimensionalidade. Em seguida, uma convolução 3×3 é utilizada para a extração de características. Depois, uma convolução 1×1 restaura a dimensionalidade original. Por fim, a entrada do bloco é somada à sua saída por meio da *skip connection*, o que facilita o fluxo de gradiente durante o treinamento.

Para integrar a rede CNN com o modelo GranSeg, algumas modificações são necessárias. Primeiramente, a camada final totalmente conectada (*fully connected*) é removida. Em seguida, a camada Conv1 é utilizada como *skip connection* 1 e conectada ao *attention gate* 1. A camada Res2 atua como *skip connection* 2 e é conectada ao *attention gate* 2. Da mesma forma, a camada Res3 é utilizada como *skip connection* 3 e conectada ao *attention gate* 3, enquanto a camada Res4 funciona como *skip connection* 4 e é conectada ao *attention gate* 4. Por fim, a camada Res5 é utilizada como *bottleneck* e conectada ao *attention gate* 5.

5.1.1.3 Atttention Gate

O bloco *Attention Gate* tem como principal objetivo aprimorar as conexões entre múltiplos *encoders* e o *decoder*, permitindo que apenas as informações mais relevantes sejam transmitidas. Esse mecanismo melhora a segmentação ao destacar regiões de interesse e suprimir ruídos, especialmente quando há múltiplas fontes de entrada. Para isso, o *Attention Gate* aplica um mecanismo de ponderação sobre as conexões de diferentes *encoders*, utilizando um mapa de atenção para modular as características antes de passá-las ao *decoder*.

Inicialmente, o bloco recebe duas entradas provenientes de diferentes *encoders*: X_1, que contém as características extraídas do primeiro *encoder*, e X_2, que representa as características extraídas do segundo *encoder*. Ambas as entradas são processadas por convoluções separadas para alinhar suas dimensões, conforme as equações:

$$Q_1 = W_1 X_1 + b_1 \tag{8}$$

$$Q_2 = W_2 X_2 + b_2 \tag{9}$$

onde W_1 e W_2 são matrizes de peso treináveis, e b_1 e b_2 são os vieses correspondentes.

Os sinais ponderados são então combinados e passam por uma ativação não linear, como *ReLU* ou *sigmoid*, para gerar um mapa de atenção α, dado por:

$$\alpha = \sigma(Q_1 + Q_2) \tag{10}$$

onde σ representa a função sigmoide, que normaliza os valores para o intervalo $[0, 1]$. O mapa de atenção α é aplicado sobre ambas as entradas, ponderando suas contribuições de acordo com:

$$X_{att} = \alpha \odot (X_1 + X_2) \tag{11}$$

onde $\odot$ representa a multiplicação elemento a elemento. Por fim, a saída X_{att} é enviada ao *decoder*, contendo apenas as informações mais relevantes para a segmentação.

A incorporação de um *Attention Gate* para combinar múltiplos *encoders* traz diversas vantagens para redes de segmentação. Primeiramente, permite a fusão inteligente de informações, combinando características aprendidas por diferentes *encoders* e aumentando a expressividade do modelo. Além disso, contribui para a redução de ruído, pois o modelo aprende a ignorar informações irrelevantes, resultando em uma segmentação mais precisa. Outra vantagem é a focalização em regiões de interesse, destacando as áreas mais importantes e aprimorando a detecção de bordas. Por fim, essa abordagem promove um aprendizado eficiente, uma vez que a combinação dos sinais melhora a reconstrução da imagem segmentada.

Assim, o Attention Gate atua como um filtro inteligente que ajusta dinamicamente a passagem de informações de múltiplos *encoders* para o *decoder*, otimizando o processo de segmentação sem adicionar um grande custo computacional.

5.1.1.4 Decoder

O *Decoder* reconstrói a imagem segmentada a partir das características extraídas pelo *encoder*. Para isso, ele realiza três funções principais. Primeiro, o *upsampling* é aplicado para aumentar a resolução espacial, utilizando camadas de convolução transposta ou outras técnicas de *upsampling*. Em seguida, ocorre a concatenação com as *skip connections*, onde as informações do *encoder*, previamente filtradas pelo bloco de atenção, são

integradas aos mapas de características *upsampled*, permitindo uma reconstrução mais detalhada. Por fim, é realizado um refinamento por meio da aplicação de novas operações de convolução, ajustando e melhorando a segmentação final.

O *Decoder* segue uma estrutura hierárquica, em que cada nível contém três operações principais. Primeiramente, ocorre o *upsampling*, realizado por meio de convoluções transpostas, que dobram a resolução espacial da entrada. Em seguida, os mapas de características extraídos pelo *Encoder* são incorporados ao processo por meio das *skip connections*, garantindo que informações espaciais de alta resolução sejam recuperadas. Por fim, novas operações de convolução são aplicadas em cada nível para refinar os mapas de características, aprimorando a segmentação e garantindo maior precisão nos contornos da imagem segmentada.

Este design permite que o modelo obtenha uma segmentação precisa, preservando os detalhes finos da imagem.

5.1.1.5 Pós-Processamento

Dada a importância da segmentação precisa para a classificação dos grãos presentes na imagem, propõe-se uma metodologia que viabilize a granulometria visual, permitindo não apenas a identificação e contagem dos grãos, mas também a extração de suas características morfológicas. Esse processo é fundamental para garantir que a segmentação preserve a individualidade de cada grão, evitando fusões ou fragmentações que possam comprometer a análise posterior.

A abordagem proposta combina técnicas de pré-processamento, segmentação e extração de características, permitindo uma análise detalhada das imagens. Essa metodologia possibilita a obtenção de informações quantitativas e estruturais essenciais para a classificação dos grãos, garantindo que a segmentação seja utilizada de forma eficaz no contexto da aplicação final.

5.1.1.6 Extração de Características

Com os grãos individualmente segmentados e rotulados, são extraídas informações quantitativas relevantes para a análise. A principal métrica considerada é a área de cada grão, expressa em unidades de pixel. Essa medida é crucial para avaliar a distribuição de tamanhos na amostra e para garantir que a segmentação preserve corretamente a morfologia dos grãos.

Além do tamanho individual de cada grão, o método também registra a contagem total de grãos na imagem. A partir dessas informações, calcula-se o tamanho médio dos grãos, fornecendo uma métrica estatística global que pode ser utilizada para comparar diferentes amostras e avaliar a variabilidade entre os experimentos.

5.1.1.7 Visualização dos Resultados

Para verificar a eficácia do processo de segmentação, os resultados são visualizados por meio da sobreposição das máscaras rotuladas às imagens originais. Cada grão segmentado recebe uma cor única, facilitando a inspeção visual da segmentação e permitindo identificar possíveis falhas, como fusões ou fragmentações incorretas.

Essa visualização não apenas auxilia na validação do método, mas também permite ajustes finos no algoritmo de segmentação, garantindo que os resultados obtidos sejam coerentes com o objetivo final de classificação dos grãos.

5.2 *Setup* e Treinamento

Para realizar o *setup* e a inicialização do modelo, devido à sua complexidade e à necessidade de treinamento intensivo, é essencial dispor de um conjunto de dados diversificado. Baseado nos experimentos detalhados no Capítulo 3 deste trabalho, propõe-se uso de um conjunto misto, composto por imagens digitais simuladas e imagens reais capturadas em ambiente controlado, complementado por imagens reais de fabrica adicionais utilizadas exclusivamente para *fine-tuning*, proporcionando o melhor aprendizado para a rede neural proposta.

Além disso, conforme observado nos resultados obtidos no Capítulo 4, o codificador *ResNet101* destacou-se ao apresentar o melhor desempenho geral, especialmente quando combinado com a saída binária. Essa configuração específica resultou na mais alta taxa de IoU (*Intersection over Union*), indicando uma capacidade superior de segmentação em relação às demais abordagens avaliadas. A utilização da saída binária mostrou-se particularmente eficaz na definição precisa dos limites dos objetos, potencializando ainda mais o desempenho do codificador *ResNet101* e confirmando esta combinação como a ideal para compor a arquitetura final proposta.

Dessa forma, a combinação do modelo U-Net com codificação dupla aprimorada por *attention gates*, a proposta de um novo codificador baseado no modelo *ResNet101* e a criação estratégica de três conjuntos de dados distintos – simulados, em ambiente controlado e provenientes de um ambiente industrial real – proporcionam um avanço significativo na tarefa de segmentação de fertilizantes. A utilização das *attention gates* melhora a capacidade do modelo em destacar regiões importantes das imagens, enquanto o codificador proposto oferece representações profundas e robustas, extraindo características detalhadas e complexas dos grãos de fertilizantes.

5.2.1 Configurações de Treinamento dos Modelos

Propõe-se uma metodologia de *setup* para testes e validação da proposta, baseada na configuração padrão descrita a seguir, garantindo a consistência e a comparabilidade dos resultados obtidos.

O treinamento deve ser realizado inicialmente por 100 épocas, utilizando uma taxa de aprendizado inicial fixada em 1×10^{-3}. Deve-se empregar o otimizador Adam durante todo o processo e adotar a Dice Loss como função de custo.

A etapa inicial de treinamento inclui 1200 imagens do ambiente virtual e 1200 imagens adquiridas em ambiente controlado. Após esse treinamento inicial, é necessário realizar uma etapa adicional de ajuste fino (*fine-tuning*), utilizando exclusivamente 200 imagens do ambiente real de fábrica, mantendo a mesma taxa de aprendizado inicial e otimizador. O ajuste fino deve ser conduzido por 30 épocas.

Para validação e acompanhamento do desempenho dos modelos durante o treinamento, deve-se utilizar um conjunto separado composto por 200 imagens originais, não incluídas nas etapas anteriores. A métrica recomendada para monitorar o desempenho dos modelos é o índice IoU.

Essa metodologia permite análises consistentes e robustas, facilitando comparações eficazes entre diferentes modelos e cenários de aplicação.

6 TESTE E VALIDAÇÃO

Este capítulo apresenta os experimentos realizados para avaliar quantitativamente e qualitativamente o desempenho do modelo GranSeg no processo de segmentação das imagens obtidas em ambiente industrial. A análise da efetividade do modelo foi realizada por meio de duas abordagens distintas: primeiramente, utilizando métricas quantitativas para avaliação direta da qualidade da segmentação; e posteriormente, aplicando um método indireto baseado na contagem dos grãos segmentados, avaliando assim a precisão qualitativa dos modelos em um contexto industrial realista.

A validação do modelo considerou sua sensibilidade em relação a diferentes perturbações do ambiente, além do impacto da disponibilidade limitada de dados para o treinamento; de forma mais precisa foram considerados os seguintes contextos:

- trepidação e variação da velocidade da esteira (com valores originais e aumentados através da aplicação de *blur*);
- presença de poeira na lente (com valores originais e aumentados através do ruído *salt and pepper*); e
- impacto da limitação do conjunto de dados disponível para o treinamento do modelo.

Nos próximos tópicos, serão detalhadas as perturbações mais relevantes e como foram implementadas, assim como os experimentos conduzidos e os resultados obtidos, ressaltando a robustez do modelo frente às adversidades do ambiente industrial.

6.1 Perturbações Relevantes

Imagens capturadas no ambiente industrial apresentam, naturalmente, uma variedade de ruídos. Embora alguns desses ruídos sejam inerentes ao contexto da indústria, certas perturbações, como variações na iluminação, podem ser atenuadas. No entanto, ruídos decorrentes da movimentação dos grãos ou da presença de poeira são mais desafiadores de mitigar sem modificações estruturais na linha de produção.

Com foco nesses ruídos, foram aplicados filtros de *blur e salt & pepper* em diferentes intensidades para degradar as imagens ao máximo criando cinco novos *datasets*. Estes experimento foram realizados com o objetivo de testar os limites do modelo, avaliando sua robustez diante de cenários extremos de degradação.

6.1.1 Aplicação de filtro Gaussian para Simulação da Perda de Qualidade das Bordas

A movimentação dos grãos na linha de produção resulta em desfoque natural nas imagens capturadas, dificultando a segmentação precisa das bordas. Para simular esse efeito, foi aplicado o filtro *gaussian* com diferentes tamanhos de *kernel*, utilizando a biblioteca CV2.

6.1.1.1 Impacto do filtro Gaussian em Diferentes Tamanhos de Kernel

A borda média dos grãos nas imagens analisadas possui entre 3 e 5 pixels de espessura. Assim, a aplicação de diferentes tamanhos de *kernel* tem os seguintes efeitos:



Figura 29: Amostra original

- Kernel (3×3): Provoca um leve desfoque, suavizando pequenas variações na textura da imagem, mas preservando a maior parte das bordas como pode ser visto na Figura 30.

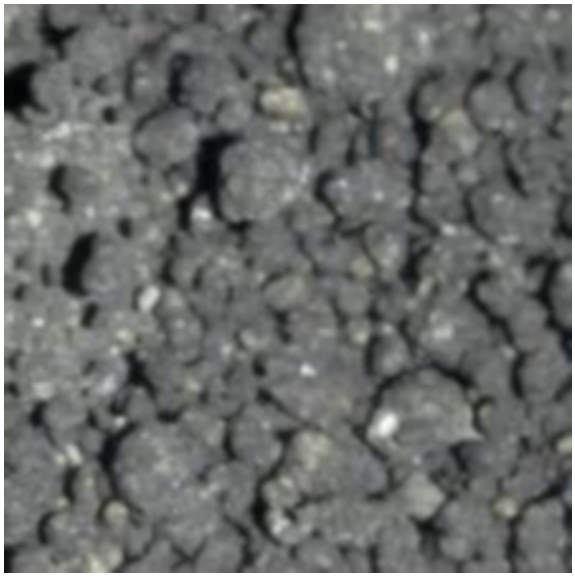


Figura 30: Amostra com blur com kernel tamanho 3x3

- Kernel (5×5): A suavização se intensifica, reduzindo a nitidez das bordas e tornando a segmentação mais desafiadora como pode ser visto na Figura 31.

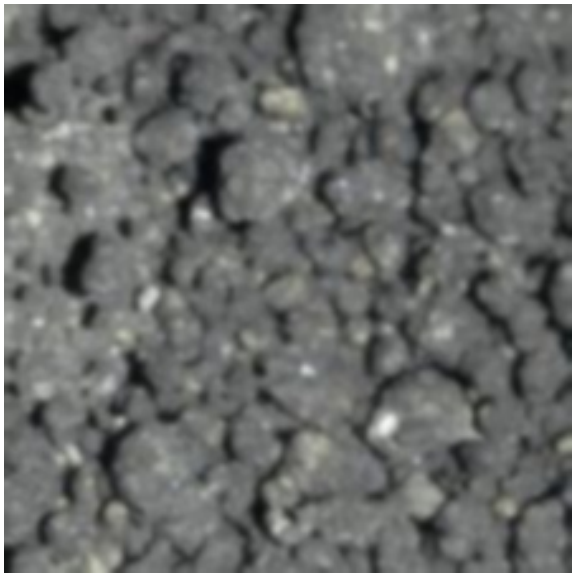


Figura 31: Amostra com blur com kernel tamanho 5x5

- Kernel (9×9): O desfoque se torna significativo, diluindo completamente detalhes finos e eliminando bordas menores que 5 pixels, o que dificulta drasticamente a detecção precisa dos contornos como pode ser visto na Figura 32.

Esse experimento visou avaliar a robustez do modelo diante da perda de definição das bordas, forçando-o a lidar com imagens em condições mais próximas da realidade industrial, onde o desfoque pode ser um fator crítico na segmentação.

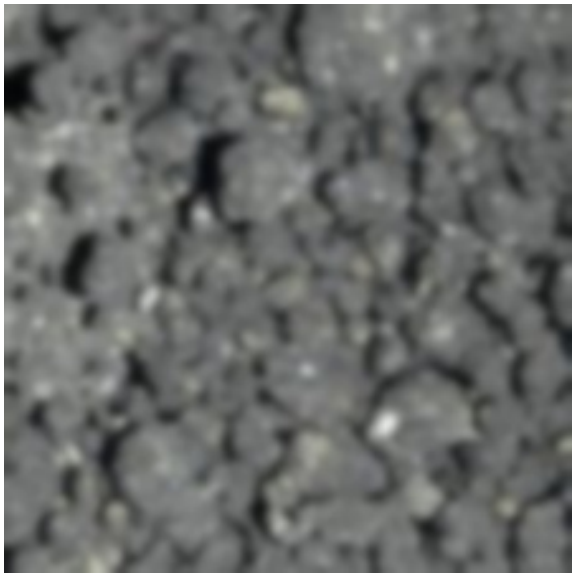


Figura 32: Amostra com blur com kernel tamanho 9x9

6.1.2 Aplicação do Ruído *Salt and Pepper*

O ruído *Salt and Pepper* foi introduzido nas imagens utilizando uma abordagem probabilística, onde cada pixel tem uma chance de ser alterado para preto (“pimenta”) ou branco (“sal”) visando simular a poeira que se acumula na lente da câmera, gerando pixels obstruídos. Esse processo foi realizado com as seguintes configurações:

- Probabilidade de 5%: Simula uma quantidade moderada de poeira, onde algumas bordas podem ser interrompidas, mas a maioria permanece identificável como pode ser visto na Figura 33.
- Probabilidade de 10%: Representa uma condição extrema de obstrução por poeira, onde uma parte significativa das bordas pode se tornar descontínua, desafiando a segmentação como pode ser visto na Figura 34.

O método aplicado segue a seguinte definição:

$$\text{img}(i,j) = \begin{cases} 0, & \text{se } r < \frac{p}{2} \quad (\text{“pimenta”}) \\ 255, & \text{se } r > 1 - \frac{p}{2} \quad (\text{“sal”}) \\ \text{img}(i,j), & \text{caso contrário} \end{cases} \tag{12}$$

onde r é um valor aleatório entre 0 e 1 para cada pixel, e p é a probabilidade definida (5% ou 10%).

6.1.2.1 Impacto do ruído salt and pepper na segmentação

A introdução desse ruído teve como objetivo avaliar a robustez do modelo frente à obstrução parcial das bordas. Os principais impactos observados são:



Figura 33: Amostra com filtro salt & pepper 5%



Figura 34: Amostra com filtro salt & pepper 10%

- discontinuidade nas bordas: O ruído pode interromper linhas de contorno dos grãos, tornando mais difícil para o modelo identificar os objetos corretamente.
- falsos contornos: A introdução de pontos brancos ou pretos aleatórios pode criar artefatos que confundem a segmentação.
- desafios para pré-processamento: Métodos tradicionais de suavização podem não ser suficientes para remover o ruído sem degradar a informação relevante.

Esse experimento visou testar a capacidade do modelo de lidar com imagens onde as bordas não são perfeitamente contínuas.

6.2 Avaliação Quantitativa da rede Granseg

Esta seção apresenta a avaliação quantitativa do modelo proposto Granseg, desenvolvido especificamente para segmentação de imagens em ambiente industrial com alto nível de sobreposição e homogeneidade de cores. O modulo de extração de bordas foi avaliado com dois algoritmos distintos: um utilizando a entrada mista RGB com segmentação obtida pelo operador determinístico de *prewitt*, e outro com entrada mista RGB associada à segmentação derivada do método Watershed.

Tambem, cinco modelos de segmentação U-net foram treinados para cada variante e avaliados utilizando a métrica val *loss Dice*, demonstrando a consistência da convergência entre as execuções.

6.2.0.0.1 Avaliação do módulo Extrator de bordas

Para a abordagem baseada em *watershed*, os valores finais de val *loss Dice* foram 0.21, 0.22, 0.23, 0.22 e 0.21, com uma média de 0.218 e um desvio padrão reduzido, indicando estabilidade nos resultados e pouca variação decorrente da inicialização aleatória dos pesos. Já para a abordagem baseada no operador de *Prewitt*, os valores obtidos foram 0.31, 0.33, 0.32, 0.35 e 0.33, com uma média de 0.328. A diferença entre as médias dos dois métodos sugere que a segmentação baseada em *watershed* obteve um desempenho significativamente superior à aquela derivada do operador *Prewitt*. Além disso, um menor desvio padrão no método *watershed* reforça sua capacidade adaptativa em diferentes execuções. Para validar a relevância estatística dessas diferenças, um teste de significância pode ser aplicado, mas, com base na magnitude das perdas, a abordagem watershed se destaca como a mais eficaz arquitetura nesse cenário industrial.

6.2.0.0.2 Estudo de ablação do módulo extrator de bordas

A fim de validar os resultados, a Granseg foi comparada com diversas arquiteturas de U-Net, incluindo a versão original e variantes com diferentes *encoders*: VGG19, DenseNet, ResNet50 e ResNet101. Para garantir uma comparação equitativa, todos os modelos

foram treinados sob as mesmas condições, utilizando dados provenientes de ambientes virtual e controlado, além de um conjunto reduzido de imagens reais do ambiente industrial, empregadas exclusivamente para ajuste fino (*fine-tuning*).

Devido à incompatibilidade entre o *ground truth* dos *datasets* desenvolvidos neste trabalho e o formato de entrada exigido pelos *datasets* do método SAM, este foi utilizado apenas para a predição. Dessa forma, não foi possível obter dados para o cálculo das métricas na avaliação quantitativa, sendo sua saída empregada exclusivamente para comparação na análise de contagem dos grãos.

A avaliação quantitativa da Granseg foi realizada com base em três métricas fundamentais: Índice de Jaccard (IoU); F-measure; e Distância de Hausdorff além das métricas, foram computados os respectivos desvios padrões, permitindo uma análise aprofundada sobre o desempenho médio e a estabilidade dos resultados. A eficiência da Granseg foi verificada em diferentes condições experimentais, abrangendo cenários adversos para testar sua eficácia.

Os resultados são apresentados nas Tabelas 6, 7, 8, 9, 10 e 11, cada uma correspondendo a um cenário específico das imagens utilizadas. O primeiro cenário refere-se às imagens originais sem degradação. O segundo contempla imagens submetidas a filtros de suavização (*blur*) para simular o borrão gerado pelo movimento dos grãos, utilizando janelas de 3×3, 5×5 e 9×9. O terceiro cenário envolve imagens com ruído impulsivo do tipo *salt and pepper*, empregado para simular a presença de poeira ocluindo a borda dos grãos, com substituição de 5% e 10% dos pixels originais.

Cada tabela apresenta os valores das métricas IoU, F-measure e Hausdorff, juntamente com seus respectivos desvios padrões (Desvio IoU, Desvio F-measure e Desvio Hausdorff). Essa abordagem detalhada permite uma análise minuciosa do desempenho da arquitetura Granseg frente às diferentes condições experimentais propostas, evidenciando sua capacidade de segmentação precisa e sua estabilidade em ambientes industriais desafiadores.

6.2.1 Análise dos Resultados

No cenário de imagens padrão, o modelo Granseg Watershed apresentou o maior IoU (0,9178) e a maior F-measure (0,9570), indicando um desempenho superior na segmentação global. O modelo VGG19-Unet obteve a menor distância de Hausdorff (2,3627), o que sugere maior precisão na delimitação dos contornos segmentados. Em relação à estabilidade, o modelo Granseg Prewitt teve a menor variação na métrica de Hausdorff (0,2678), enquanto a U-Net original apresentou a maior variabilidade (desvio padrão de 0,6051). Esses resultados podem ser observados na Tabela 6. O método Granseg Watershed obteve resultados satisfatórios na segmentação em condições reais de fábrica, indicando adequação para aplicações práticas em ambientes industriais.

Rede	IoU	F-measure	Hausdorf	Desvio IoU	Desvio F-measure	Desvio Hausdorf
U-net	0.859092	0.924080	2.584585	0.0201	0.011606	0.605168
Vgg19-Unet	0.865682	0.927869	2.362727	0.021184	0.012107	0.285898
Dense-Unet	0.862442	0.926004	2.375982	0.0210	0.0120	0.287975
Resnet50-Unet	0.863475	0.926593	2.378203	0.021625	0.012396	0.292767
Resnet101-Unet	0.86426	0.927047	2.380282	0.021501	0.012317	0.29487
Granseg Prewitt	0.862956	0.926303	2.354669	0.020896	0.011981	0.267814
Granseg Watershed	0.917833	0.957036	2.428826	0.020503	0.011233	0.305046

Tabela 6: Tabela com a comparação de diferentes redes com entrada padrão.

Ao aplicar um filtro de desfoque 3×3, o Granseg Watershed manteve o melhor desempenho global, com IoU de 0,8983 e F-measure de 0,9463, mas com uma leve redução em relação ao cenário sem suavização. O VGG19-Unet continuou sendo o modelo mais preciso na detecção dos contornos, com distância de Hausdorff de 2,3731. O modelo U-Net original apresentou a maior variação nos resultados (desvio padrão de 0,6450), enquanto o Granseg Prewitt manteve a maior estabilidade (desvio padrão de 0,2742). Esses valores são detalhados na Tabela 7.

Mesmo sob a aplicação de um filtro de desfoque (blur 3x3), o método Granseg Watershed demonstrou capacidade de preservar detalhes importantes na segmentação das imagens, indicando sua adequação para lidar com situações em que há perda parcial de nitidez. Esse comportamento sugere que a abordagem utilizada pelo método é eficiente na identificação correta dos contornos das partículas, mantendo a qualidade da segmentação mesmo em condições de degradação da imagem, algo frequentemente encontrado em cenários industriais reais.

Rede	IoU	F-measure	Hausdorf	Desvio IoU	Desvio F-measure	Desvio Hausdorf
U-net	0.859175	0.924123	2.579105	0.020558	0.011832	0.644977
Vgg19-Unet	0.864431	0.927142	2.373144	0.021768	0.012465	0.288233
Dense-Unet	0.862856	0.926225	2.382301	0.022457	0.012885	0.297975
Resnet50-Unet	0.863079	0.926356	2.382301	0.022275	0.012777	0.291324
Resnet101-Unet	0.863276	0.926471	2.379141	0.022156	0.012709	0.289672
Granseg Prewitt	0.862370	0.925958	2.365233	0.021479	0.012324	0.274241
Granseg Watershed	0.898296	0.946276	2.422793	0.022501	0.012506	0.304942

Tabela 7: Tabela com a comparação de diferente redes com entrada com blur 3x3

Com o filtro 5×5, a perda de definição nas imagens impactou ainda mais o desempenho geral dos modelos. O Granseg Watershed obteve IoU de 0,8796 e F-measure de 0,9358, evidenciando um declínio progressivo na segmentação. O VGG19-Unet manteve a menor distância de Hausdorff (2,3817), seguido pelos modelos Dense-Unet e ResNet. O modelo U-Net original continuou com a maior variação nos resultados (desvio padrão de 0,6630), enquanto o Granseg Prewitt permaneceu como o mais estável (desvio padrão de 0,2742). Esses resultados estão organizados na Tabela 8.

Esses resultados reforçam a importância das estratégias empregadas pelo Granseg Watershed e Granseg Prewitt, cujos desempenhos destacam a relevância de métodos que exploram diretamente as características morfológicas e estruturais das imagens. Esse comportamento é particularmente valioso em ambientes industriais reais, onde imagens frequentemente apresentam variações significativas na qualidade e clareza dos detalhes, exigindo modelos que sejam capazes de se adaptar a diferentes graus de degradação visual sem comprometimento crítico da segmentação.

Rede	IoU	F-measure	Hausdorf	Desvio IoU	Desvio F-measure	Desvio Hausdorf
U-net	0.859291	0.924188	2.576627	0.020778	0.011958	0.662948
Vgg19-Unet	0.863403	0.926543	2.381671	0.022205	0.012732	0.291896
Dense-Unet	0.862590	0.926071	2.382301	0.022477	0.012899	0.291324
Resnet50-Unet	0.862759	0.926169	2.382301	0.022473	0.012895	0.291324
Resnet101-Unet	0.862658	0.926111	2.382301	0.022423	0.012872	0.291324
Granseg Prewitt	0.861609	0.925512	2.378477	0.021935	0.012600	0.290527
Granseg Watershed	0.879643	0.935795	2.436801	0.024036	0.013585	0.307005

Tabela 8: Tabela com a comparação de diferente redes com entrada com blur 5x5

Na análise com o filtro de desfoque 9×9, o modelo VGG19-Unet passou a apresentar o maior IoU (0,8626) e a maior F-measure (0,9261), superando o Granseg Watershed, que obteve um desempenho inferior (IoU de 0,8559). O VGG19-Unet também manteve a menor distância de Hausdorff (2,3823). O modelo U-Net original apresentou a maior variação nos resultados (desvio padrão de 0,7446), indicando instabilidade com imagens altamente degradadas. Modelos com encoders pré-treinados, como VGG19, DenseNet e ResNet, demonstraram menor perda de desempenho em comparação ao Granseg Watershed. Os detalhes desses resultados podem ser encontrados na Tabela 9.

Esse resultado demonstra que, em situações extremas de degradação visual (blur 9×9), as redes com encoders pré-treinados, especialmente o VGG19-Unet, apresentam maior resiliência e capacidade de adaptação a imagens altamente desfocadas em comparação ao método Granseg Watershed. Essa mudança no desempenho sugere que, sob alta perda de detalhes visuais, estratégias baseadas em aprendizado prévio a partir de grandes conjuntos de imagens, como ImageNet, podem oferecer vantagens adicionais na preservação da qualidade da segmentação. Por outro lado, modelos baseados em métodos mais específicos de detecção de contornos, como o Granseg Watershed, podem ter seu desempenho comprometido pela significativa suavização dos detalhes críticos para sua abordagem.

Rede	IoU	F-measure	Hausdorf	Desvio IoU	Desvio F-measure	Desvio Hausdorf
U-net	0.859915	0.924545	2.608017	0.021132	0.012156	0.744577
Vgg19-Unet	0.862592	0.926072	2.382301	0.022488	0.012906	0.291324
Dense-Unet	0.862567	0.926058	2.382301	0.022473	0.012897	0.291324
Resnet50-Unet	0.862576	0.926063	2.382301	0.022473	0.012897	0.291324
Resnet101-Unet	0.861990	0.925724	2.382301	0.022574	0.012968	0.291324
Granseg Prewitt	0.861107	0.925216	2.386375	0.022395	0.012874	0.294583
Granseg Watershed	0.855866	0.922156	2.402124	0.024104	0.013926	0.297827

Tabela 9: Tabela com a comparação de diferente redes com entrada com blur 9x9

6.2.2 Considerações finais quanto a sensibilidade do Método a velocidade de esteira

Os experimentos mostraram que, em imagens sem suavização, o modelo Granseg Watershed apresentou o melhor desempenho global, com os maiores valores de IoU e F-measure. Entretanto, conforme a degradação da imagem aumentou, esse modelo perdeu eficiência, indicando maior sensibilidade à perda de detalhes. Em contrapartida, os modelos baseados em encoders pré-treinados, principalmente o VGG19-Unet, mantiveram um desempenho mais consistente, com menor variação na precisão dos contornos segmentados.

A análise da métrica de Hausdorff mostrou que o VGG19-Unet teve a menor distância média em todos os cenários, o que indica maior precisão na segmentação dos limites dos objetos. Em relação à estabilidade dos resultados, o Granseg Prewitt apresentou os menores desvios-padrão para a métrica de Hausdorff, sugerindo maior consistência em diferentes condições de entrada.

Os resultados indicam que o Granseg Watershed é a melhor escolha para imagens sem degradação, mas sua queda de desempenho com o aumento da suavização pode limitar sua aplicação em cenários onde há perda de definição na captura das imagens. Já os modelos baseados em encoders pré-treinados, como VGG19, DenseNet e ResNet, demonstraram melhor adaptação a degradações moderadas e severas, sendo opções mais confiáveis quando a qualidade das imagens não pode ser garantida.

A escolha do modelo depende do nível de degradação das imagens na aplicação final. Em ambientes onde as imagens possuem boa definição, o Granseg Watershed apresenta os melhores resultados. Em situações onde as imagens podem sofrer perda de detalhes, os modelos com encoders pré-treinados oferecem um desempenho mais previsível e estável. Esses resultados destacam a necessidade de avaliar não apenas a precisão global, mas também a estabilidade dos modelos para aplicações industriais.

6.2.3 Impacto do Ruído Salt and Pepper nos Modelos de Segmentação

O ruído impulsivo do tipo *salt and pepper* é um fenômeno comum em ambientes industriais, onde poeira e outras impurezas podem grudar na lente da câmera e obstruir a visão de certos pixels, ocultando partes das bordas dos grãos e dificultando a segmentação

precisa. Esse ruído afeta aleatoriamente os pixels da imagem, transformando alguns em valores máximos (branco) e outros em valores mínimos (preto), simulando condições reais de degradação.

Nesta subseção, são apresentados os resultados obtidos pelos modelos de segmentação ao serem expostos a diferentes níveis desse ruído, com taxas de 5% e 10%. As Tabelas 10 e 11 apresentam a comparação de diferentes modelos considerando as métricas IoU, F-measure e distância de Hausdorff, além dos respectivos desvios padrões.

Rede	IoU	F-measure	Hausdorf	Desvio IoU	Desvio F-measure	Desvio Hausdorf
U-net	0.858389	0.923674	2.587384	0.020159	0.011613	0.613008
Vgg19-Unet	0.863370	0.926535	2.365887	0.021477	0.012311	0.287752
Dense-Unet	0.862654	0.926117	2.375353	0.021825	0.012520	0.288539
Resnet50-Unet	0.850297	0.918932	2.385425	0.022563	0.013228	0.293219
Resnet101-Unet	0.862518	0.926052	2.523973	0.020837	0.011956	0.519888
Granseg Prewitt	0.853743	0.920953	2.358558	0.021801	0.012641	0.263415
Granseg Watershed	0.817431	0.899394	2.842470	0.021441	0.012906	0.279116

Tabela 10: Tabela com a comparação de diferente redes com entrada com 5% salt and pepper

Rede	IoU	F-measure	Hausdorf	Desvio IoU	Desvio F-measure	Desvio Hausdorf
U-net	0.857049	0.922894	2.569549	0.020320	0.011727	0.611896
Vgg19-Unet	0.860954	0.925135	2.375673	0.021848	0.012558	0.290520
Dense-Unet	0.860676	0.924977	2.381671	0.021688	0.012472	0.291896
Resnet50-Unet	0.835217	0.909977	2.389017	0.026695	0.016109	0.295801
Resnet101-Unet	0.861506	0.925465	2.471598	0.021006	0.012066	0.407815
Granseg Prewitt	0.845769	0.916281	2.373102	0.022463	0.013151	0.279156
Granseg Watershed	0.792668	0.884196	3.067903	0.020752	0.012858	0.271105

Tabela 11: Tabela com a comparação de diferente redes com entrada com 10% salt and pepper

Os resultados indicam que, sob ruído de 5%, o modelo VGG19-Unet apresentou o melhor desempenho geral, com IoU de 0,8634 e F-measure de 0,9265, seguido de perto pelo Dense-Unet. O modelo Granseg Watershed, por outro lado, sofreu uma queda expressiva no desempenho (IoU de 0,8174 e F-measure de 0,8994), sugerindo alta vulnerabilidade a ruídos impulsivos. Em termos de precisão local, a distância de Hausdorff foi menor no VGG19-Unet (2,3659), enquanto o Granseg Watershed obteve o pior resultado (2,8425). A estabilidade dos resultados também favoreceu o VGG19-Unet, que apresentou o menor desvio padrão na distância de Hausdorff (0,2878), enquanto a U-Net original apresentou maior instabilidade (0,6130).

Com um aumento para 10% de ruído, a tendência observada se manteve: o VGG19-Unet continuou sendo o modelo mais robusto, com IoU de 0,8609 e F-measure de 0,9251, enquanto o Granseg Watershed sofreu um impacto ainda mais significativo (IoU de 0,7927 e F-measure de 0,8842). A distância de Hausdorff aumentou para todos os modelos,

sendo novamente menor no VGG19-Unet (2,3757) e mais elevada no Granseg Watershed (3,0679). A estabilidade seguiu o padrão anterior, com o VGG19-Unet apresentando um desvio padrão reduzido (0,2905) e a U-Net original mantendo alta variabilidade (0,6119).

Como complemento, é importante ressaltar que o impacto significativo do ruído impulsivo sobre o desempenho do método Granseg Watershed pode ser atribuído à sua dependência direta da continuidade das bordas na segmentação das partículas. A presença desse tipo específico de ruído, que altera drasticamente a intensidade dos pixels em pontos isolados, interfere na precisão dos algoritmos baseados em gradientes ou variações locais, resultando em uma segmentação menos precisa. Em contraste, as redes U-Net com encoders pré-treinados, como o VGG19-Unet, utilizam representações mais generalizadas aprend

6.2.4 Considerações finais quanto a sensibilidade do Método a Poeira

Embora o método Granseg Watershed tenha apresentado sensibilidade ao ruído impulsivo, é importante destacar suas qualidades positivas observadas ao longo dos experimentos. Em condições de baixa degradação, como nas imagens originais e com baixo nível de desfoque, esse método obteve consistentemente os melhores resultados de segmentação, especialmente em relação à métrica IoU e F-measure, indicando alta capacidade de reconhecer com precisão as estruturas das partículas. Essa vantagem inicial reforça a eficiência do método em ambientes controlados ou cenários industriais em que as imagens apresentam níveis moderados de interferência visual. Além disso, a abordagem determinística do Granseg Watershed pode ser particularmente vantajosa em situações onde recursos computacionais limitados ou a necessidade de interpretabilidade dos resultados sejam prioridades.

6.3 Avaliação da Metodologia de Contagem dos Grãos

Dada a importância da segmentação precisa para a classificação e granulometria dos grãos presentes nas imagens, foi aplicada a metodologia proposta que permite não apenas identificar e contar os grãos segmentados, mas também avaliar o tamanho individual de cada grão identificado. Esse tipo de abordagem é essencial, uma vez que fornece informações quantitativas adicionais, fundamentais para análises granulométricas mais detalhadas no contexto industrial.

O método empregado combina técnicas avançadas de pré-processamento das imagens, segmentação precisa dos contornos dos grãos e posterior extração de características, permitindo assim não apenas uma contagem robusta dos grãos identificados, mas também uma análise detalhada sobre suas dimensões morfológicas. Tal abordagem é particularmente relevante, visto que erros na etapa de segmentação podem impactar significativamente as análises posteriores, como a classificação ou estimativa quantitativa.

Nesta seção são apresentados os resultados obtidos por meio desta metodologia, permitindo avaliar o desempenho dos modelos analisados anteriormente quanto à capacidade de segmentação precisa, contando e identificando os grãos presentes nas imagens. Além disso, alguns resultados obtidos são comparados ao desempenho obtido por um modelo do estado da arte em segmentação, o *Segment Anything Model* (SAM), permitindo uma visão mais clara sobre o posicionamento dos modelos propostos em relação às técnicas mais avançadas atualmente disponíveis na literatura.



Figura 35: Exemplo de entrada em rgb

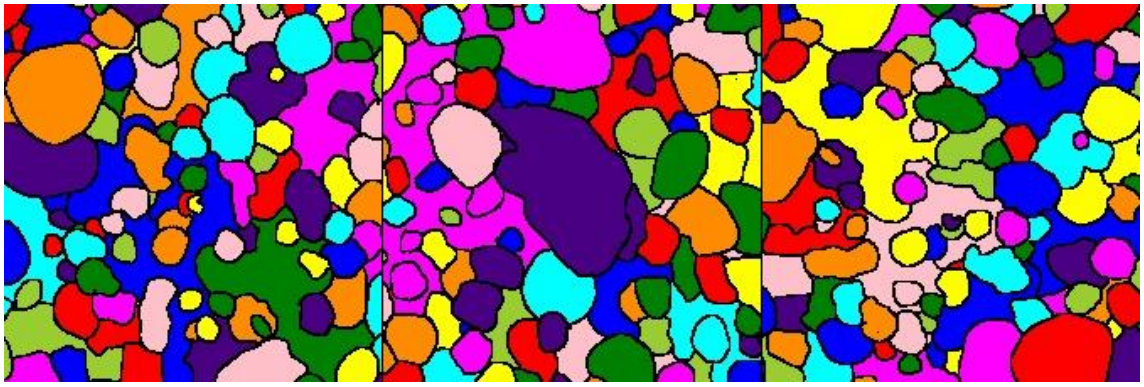


Figura 36: Exemplo de saida esperada

6.3.1 Resultados Iniciais Sem Aplicação de Ruído

Nesta subseção, são apresentadas as análises iniciais realizadas utilizando imagens RGB sem nenhuma aplicação adicional de ruído. São comparados os resultados obtidos a partir de quatro métodos distintos: o modelo baseado em VGG19, que demonstrou bom desempenho geral; o modelo DenseNet, que obteve resultados especialmente positivos com dados provenientes de ambientes controlados; o método GranSeg utilizando Watershed, que apresentou o melhor desempenho medido pelo índice IoU em imagens originais capturadas diretamente em ambiente industrial; e, finalmente, o modelo SAM, empregado como referência atual do estado da arte em segmentação.

As Figuras 35, 36, mostram a entrada alimentada nas redes e o resultado esperado.

A Figura 37 apresenta, em conjunto, os resultados obtidos pelos modelos Granseg Watershed, DenseNet, VGG19-Unet e SAM para as três imagens analisadas, permitindo uma comparação visual direta entre os métodos.

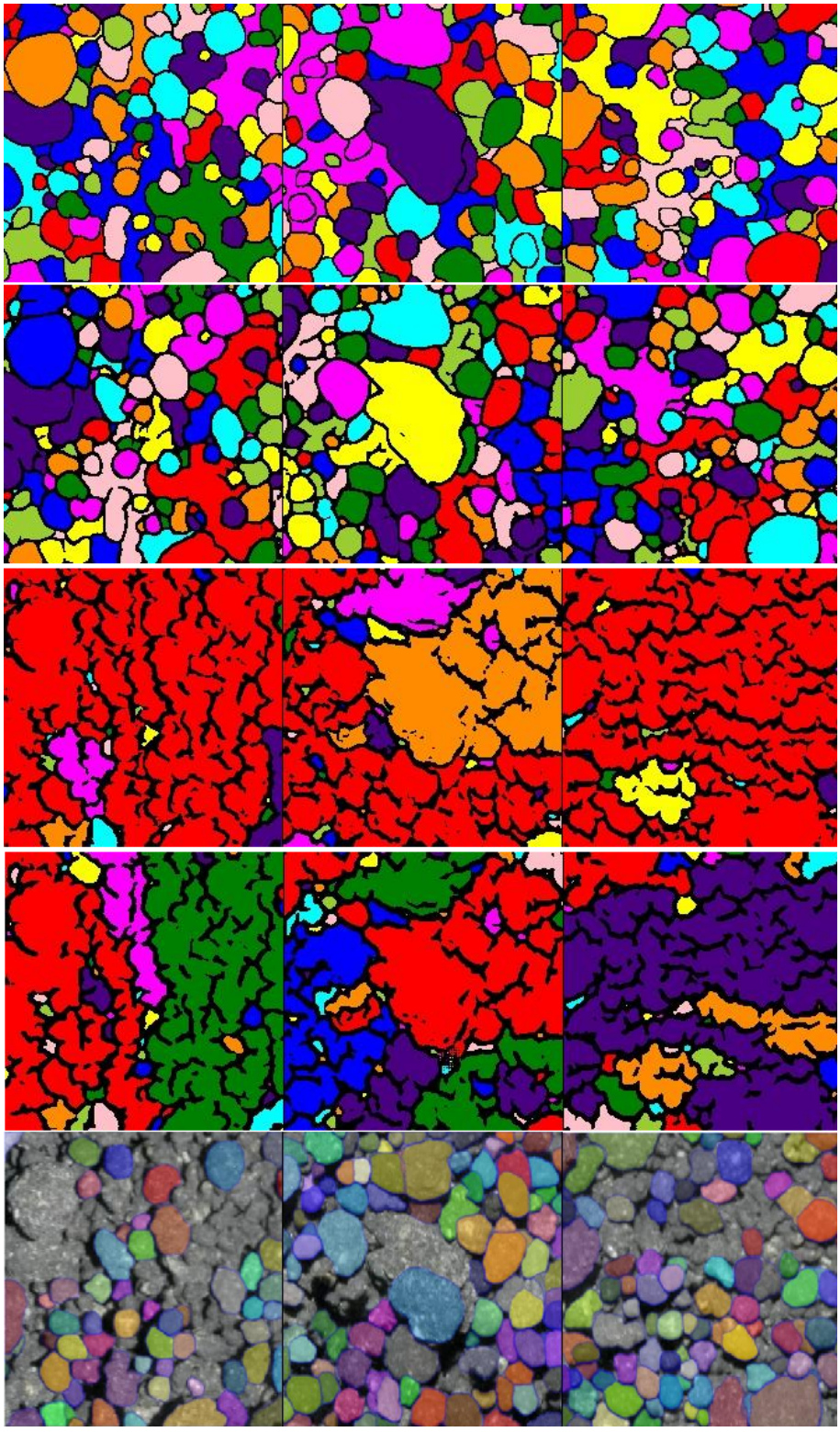


Figura 37: Comparação visual dos resultados obtidos por diferentes métodos de segmentação sem a presença de ruído nas imagens originais. A primeira linha mostra a máscara esperada (ground truth), seguida pelas saídas dos métodos: Granseg Watershed (segunda linha), DenseNet (terceira linha), VGG19-Unet (quarta linha) e Segment Anything Model (SAM, quinta linha). É possível observar visualmente que o método Granseg Watershed apresenta segmentações com maior integridade estrutural e continuidade nas bordas dos objetos segmentados, em comparação às segmentações mais fragmentadas geradas pelos demais métodos.

É possível notar que o modelo GranSeg Watershed, ilustrado na segunda linha da Figura 37, obteve uma segmentação ideal, com linhas bem definidas que permitiram à metodologia de análise dos grãos colorir cada grão de forma similar à máscara esperada, correspondendo às expectativas levantadas pelo alto valor de IoU apresentado previamente (0.91).

Já a saída do modelo DenseNet, apresentada na terceira linha da Figura 37, embora previamente no capitulo 4 tenha demonstrado boa segmentação para o dataset em ambiente controlado, não foi capaz de segmentar adequadamente uma imagem proveniente do dataset em ambiente de fábrica, falhando na criação de linhas contínuas necessárias para uma segmentação ideal.

Da mesma forma, a saída da rede VGG19, representada na quarta linha da Figura 37, também demonstrou dificuldades em manter linhas contínuas nas bordas dos grãos, comprometendo uma análise precisa dos mesmos.

Por fim, o método SAM, exibido na ultima linha da Figura 37, apresentou uma segmentação ideal em diversos grãos, porém encontrou dificuldades em identificar corretamente grãos sobrepostos, não sendo capaz de segmentar adequadamente todos os grãos presentes na imagem.

6.3.2 Resultados com Aplicação de Desfoque (Blur 3x3)

Nesta subseção, são avaliados os impactos da aplicação de desfoque com uma janela 3x3 nas imagens de entrada. Os resultados dos quatro métodos anteriormente descritos são novamente apresentados para análise comparativa. Essa abordagem permite avaliar a robustez dos modelos frente a pequenas alterações e perda de nitidez das imagens originais.

A Figura 39 apresenta os resultados obtidos pelos modelos Granseg Watershed, DenseNet, VGG19-Unet e SAM para as três imagens analisadas após a aplicação do filtro de desfoque (blur 3×3 fig 38), permitindo uma comparação visual direta do impacto dessa degradação nos diferentes métodos.



Figura 38: Entrada rgb com filtro blur 3x3

A saída da rede GranSeg Watershed (segunda linha da Figura 39) manteve-se com um alto índice de segmentação de grãos, apresentando apenas pequenas perdas na

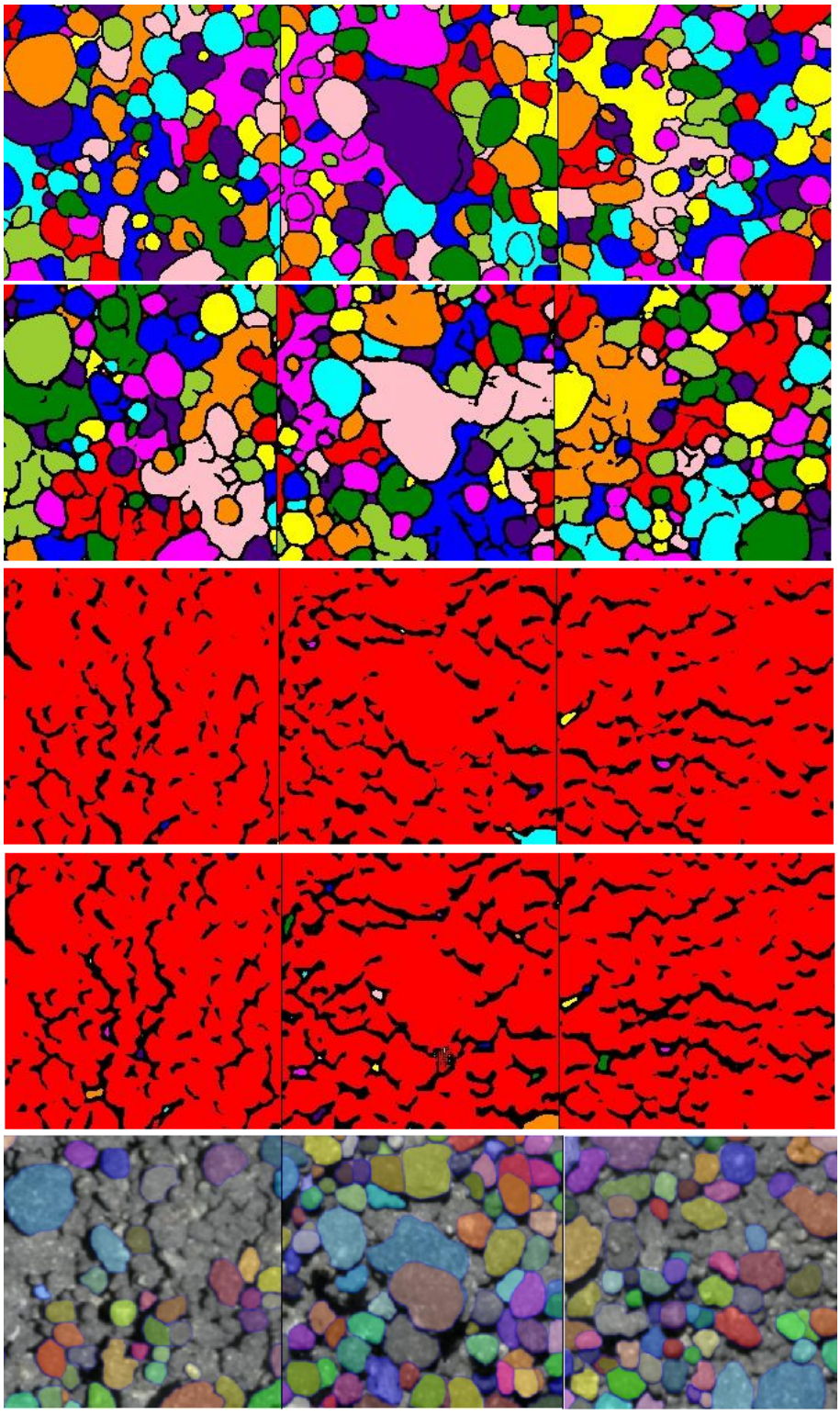


Figura 39: Comparação visual dos resultados obtidos pelos diferentes modelos após a aplicação do filtro de desfoque (blur 3×3). A primeira linha apresenta a máscara esperada (ground truth), seguida pelas saídas dos métodos Granseg Watershed (segunda linha), DenseNet (terceira linha), VGG19-Unet (quarta linha) e Segment Anything Model (SAM, quinta linha). Observa-se visualmente o impacto do desfoque nas segmentações, sendo possível avaliar qualitativamente a resistência dos métodos em manter a integridade das bordas e estruturas segmentadas.



Figura 40: Imagens de entrada com filtro de blur de 9x9

identificação de alguns grãos.

Por outro lado, as saídas dos métodos DenseNet (terceira linha da Figura 39) e VGG19 U-net (quarta linha da Figura 39) perderam totalmente a continuidade das linhas de segmentação, causando uma falha no método de análise de grãos.

Por fim, o método SAM (ultima linha da Figura 39) permaneceu consistente na segmentação das imagens, mesmo sob a influência do filtro de blur 3x3, sofrendo apenas uma leve perda na identificação dos grãos menores.

6.3.3 Resultados com Aplicação de Desfoque Intenso (Blur 9x9)

Nesta subseção, são discutidos os efeitos causados por um desfoque mais significativo, utilizando uma janela de 9x9 nas imagens originais. Essa avaliação possibilita identificar como os diferentes modelos se comportam em situações com perda considerável de detalhes das bordas dos grãos.

As Figuras 40, 41, mostram a entrada alimentada nas redes e o resultado esperado.

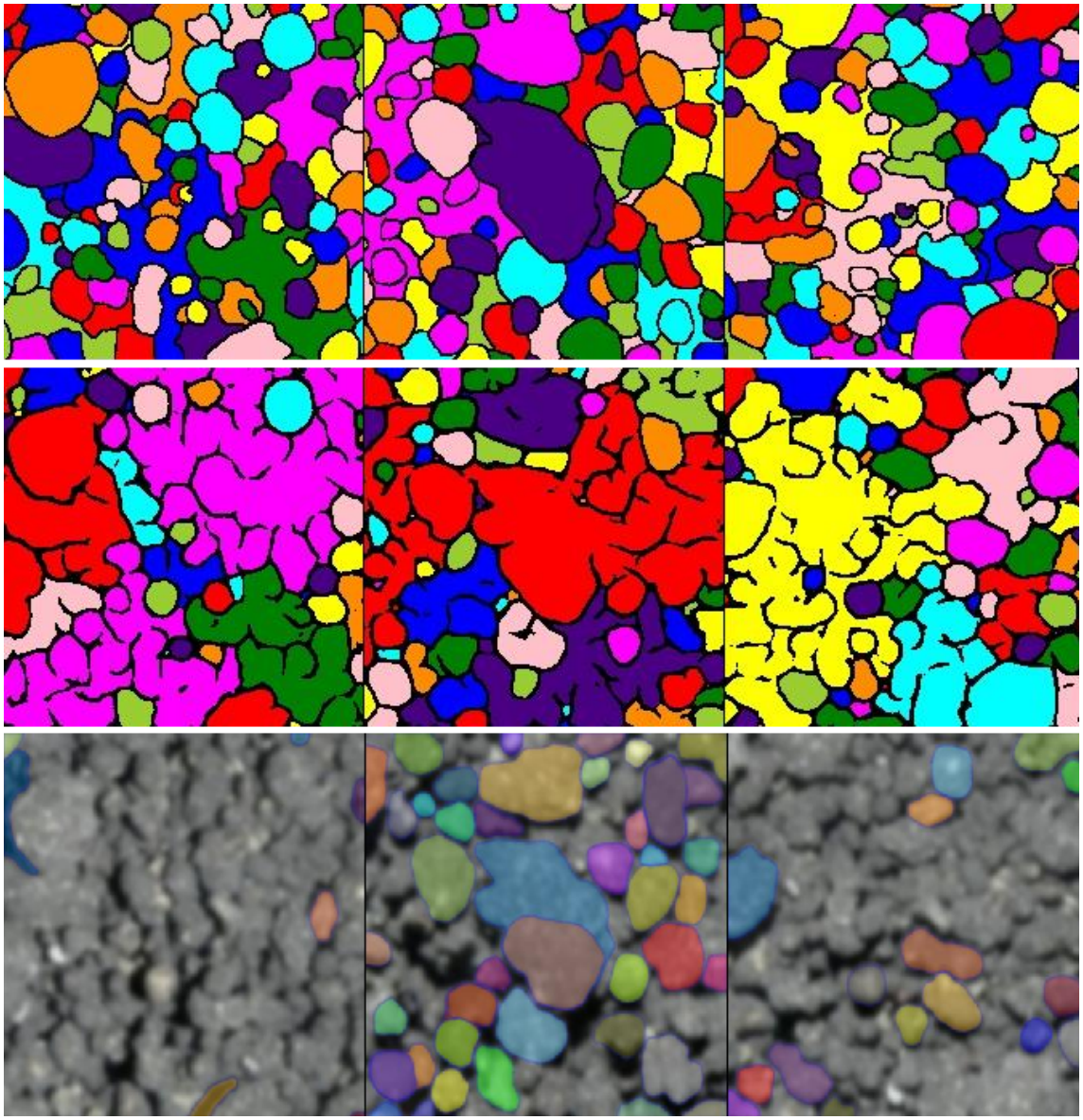


Figura 41: Resultados visuais obtidos após a aplicação do filtro de desfoque mais intenso (blur 9×9). A primeira linha mostra as máscaras esperadas (ground truth), enquanto a segunda e terceira linhas apresentam as segmentações obtidas pelo método Granseg Watershed e pelo Segment Anything Model (SAM), respectivamente. Nota-se o efeito significativo do desfoque sobre os resultados, permitindo avaliar qualitativamente a capacidade dos métodos em manter a coerência estrutural e a continuidade das bordas segmentadas sob alta degradação visual.

Devido ao forte efeito atenuador do filtro blur com janela 9×9, nenhum dos métodos avaliados obteve resultados satisfatórios. No caso do método GranSeg (segunda linha da Figura 41), a maioria dos grãos maiores apresentou falhas significativas na manutenção de uma borda contínua, dificultando sua identificação. No entanto, um refinamento no pós-processamento poderia potencialmente aumentar a quantidade de grãos corretamente identificados. Para os grãos menores, a segmentação parcial permitiu sua identificação, viabilizando a análise após o descarte daqueles com dimensões muito superiores ao espe-

rado.

No caso do método SAM (ultima linha da Figura 41), os resultados foram altamente inconsistentes, com falhas totais em segmentação em algumas imagens e segmentações razoáveis em outras.

Este resultado era esperado, visto que a janela utilizada no filtro de blur (9x9) excedia significativamente a espessura média das bordas dos grãos presentes nas imagens.

6.3.4 Resultados com Aplicação de Ruído Salt and Pepper (5%)

Por fim, esta subseção aborda os resultados da aplicação de ruído do tipo Salt and Pepper com uma probabilidade de 5%. A análise desses resultados permite uma compreensão adicional sobre a capacidade dos métodos avaliados em resistir a ruídos aleatórios comuns em ambientes industriais reais.

As Figuras 42, 36, mostram a entrada alimentada nas redes e o resultado esperado.



Figura 42: Entrada RGB com 5% de chance de salt & pepper

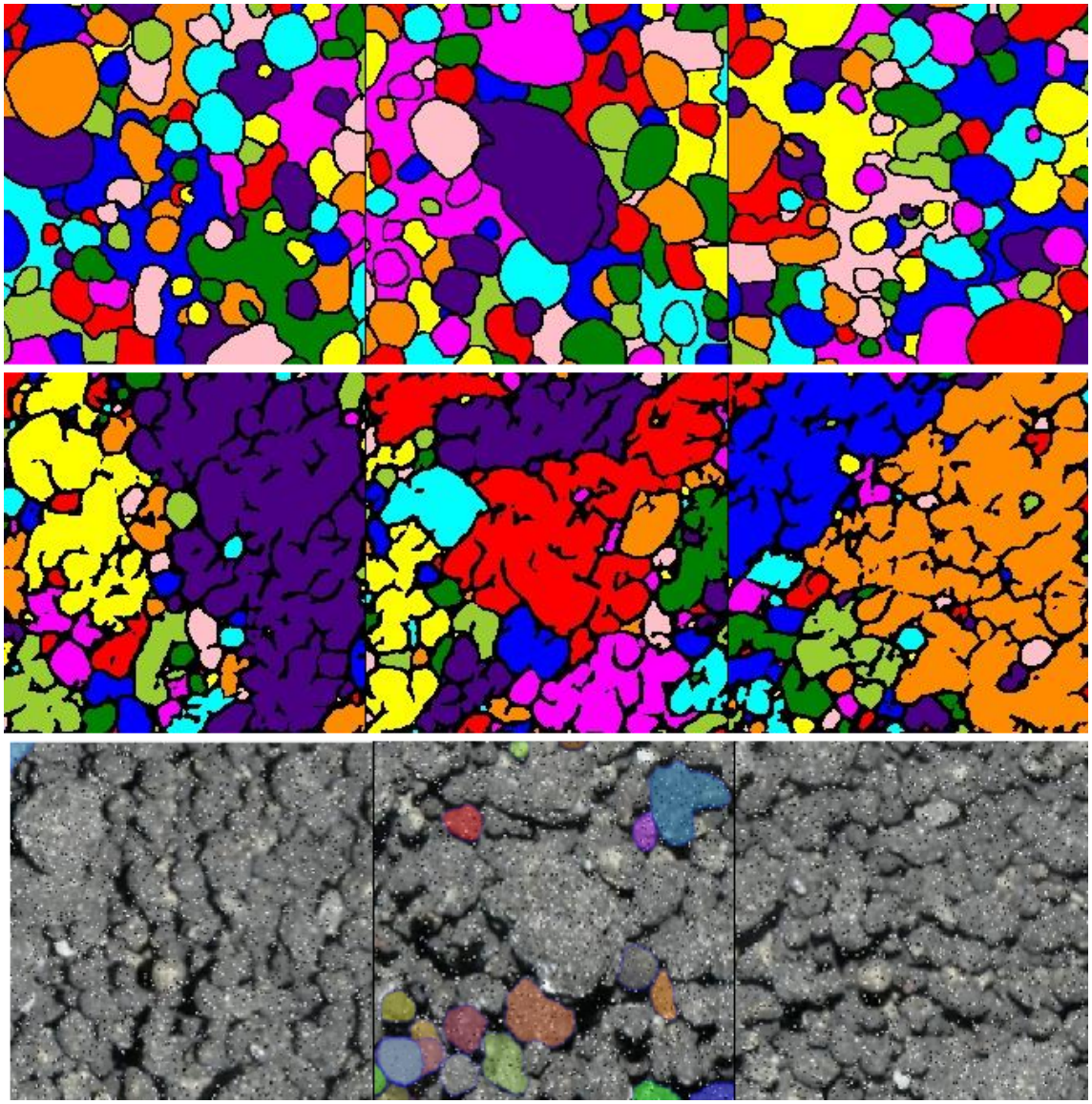


Figura 43: Resultados visuais dos modelos submetidos ao ruído impulsivo do tipo *salt and pepper* com intensidade de 5%. A primeira linha apresenta as máscaras esperadas (ground truth), enquanto as linhas seguintes exibem as segmentações obtidas pelos métodos Granseg Watershed (segunda linha) e Segment Anything Model (SAM, terceira linha). Observa-se qualitativamente o impacto desse ruído na capacidade dos métodos em preservar a continuidade e integridade das estruturas segmentadas.

Os experimentos com a aplicação de ruído do tipo Salt and Pepper com 5% de probabilidade apresentaram a maior perda na segmentação dos grãos, devido à interferência desse tipo de ruído na definição das bordas.

Para o método GranSeg com Watershed (segunda linha da Figura 43), a maioria dos grãos maiores falhou em manter a continuidade da borda, enquanto apenas os grãos menores foram propriamente segmentados.

Já o método SAM (terceira linha da Figura 43) falhou em identificar corretamente um único grão na maioria das imagens analisadas.

Embora os resultados obtidos pelo método GranSeg ainda possam ser utilizados após

a aplicação de um threshold de tamanho para remover os grãos de dimensões excessivas, fica evidente que um dos maiores desafios a serem superados na segmentação está na presença de partículas de poeira e ruídos que comprometem a continuidade das bordas.

6.3.5 Considerações

Dado os resultados apresentados, embora a análise de IoU previamente mencionada não corrobore para todos os casos, o método de segmentação GranSeg tem consistentemente retornado imagens com segmentação satisfatória para pequenos grãos, mesmo quando há sobreposição. Além disso, em condições ideais, também apresenta uma segmentação adequada para grãos maiores.

O método de estado da arte SAM, apesar de oferecer uma segmentação de alto nível, enfrenta dificuldades na segmentação de um alto número de grãos agregados com grande sobreposição, especialmente quando os grãos possuem maior tamanho. Além disso, o método falhou completamente quando havia a presença de poeira e apresentou um desempenho altamente inconsistente quando o filtro de blur 9x9 foi aplicado.

7 CONCLUSÃO

Este trabalho teve como objetivo principal desenvolver um método não invasivo para análise granulométrica de fertilizantes por meio de técnicas avançadas de segmentação de imagens, utilizando uma visão superior da linha de produção. Durante a pesquisa, destacou-se a necessidade de compreender profundamente as características e ruídos das imagens obtidas nesse contexto, especialmente devido à irregularidade das formas dos grãos, à homogeneidade das cores e à ausência de bordas bem definidas.

Esses desafios inviabilizaram o uso isolado de métodos clássicos de segmentação, exigindo soluções inovadoras. Assim, foi proposto o desenvolvimento de uma abordagem baseada em aprendizado profundo (Granseg), fundamentada em uma arquitetura U-net adaptada, com múltiplos encoders e atenção nas conexões intermediárias (attention gates), utilizando uma proposta adaptatada do modelo ResNet101 como encoder principal. Além disso, o modelo Granseg utiliza duas entradas simultâneas: uma imagem original no formato RGB e uma imagem pré-segmentada por métodos clássicos (como Prewitt e Watershed), o que facilita e guia a convergência do treinamento. A saída gerada é uma imagem binária contendo as bordas dos grãos, permitindo a realização da análise granulométrica.

A Granseg demonstrou desempenho superior ao modelo de referência (Segment Anything Model - SAM), principalmente devido à sua especialização na segmentação de objetos homogêneos e sobrepostos, além da eficiência em treinamento com datasets reduzidos.

Dado que não existiam datasets públicos adequados, foi proposta uma metodologia para criação e uso de datasets baseada em três conj untos específicos: (i) um dataset sintético gerado por simulação computacional; (ii) um dataset obtido em ambiente controlado, com iluminação homogênea e protocolos definidos; e (iii) um dataset capturado diretamente em ambiente industrial, com condições reais de produção. A combinação estratégica desses datasets revelou que o uso inicial das imagens sintéticas e controladas, seguido por um fine-tuning com um número limitado de imagens reais (apenas 200), proporcionou melhores resultados na segmentação em ambientes industriais, validando a eficácia dessa metodologia híbrida de treinamento.

Em síntese, o método Granseg cumpriu satisfatoriamente o objetivo proposto, demonstrando potencial para automatizar e melhorar significativamente a análise granulométrica em contextos industriais reais. As principais contribuições deste trabalho incluem:

- desenvolvimento e disponibilização de metodologia para criação e uso de datasets diversificados (sintético, controlado e real), essenciais para treinamento eficaz em cenários industriais;
- análise aprofundada sobre o impacto positivo do pré-treinamento com dados sintéticos e controlados, permitindo reduzir a dependência de grandes quantidades de dados reais;
- proposição de um método robusto e eficiente, especializado em destacar bordas em imagens com alta sobreposição e ruído; e
- demonstração prática da aplicabilidade industrial do modelo proposto, com potencial de integração em sistemas automatizados de visão computacional.

Como trabalhos futuros, sugere-se explorar a distribuição granulométrica dos fertilizantes em esteiras, especialmente para entender como ocorre a segregação dos grãos por tamanho ao longo da profundidade.

Adicionalmente, recomenda-se desenvolver um método automatizado de calibração que permita determinar precisamente a relação entre a distância real e a dimensão em pixels das imagens, facilitando assim a aplicação prática do método desenvolvido em ambientes industriais variados.

REFERÊNCIAS

[Abdulateef e Salman 2021]ABDULATEEF, S.; SALMAN, M. A comprehensive review of image segmentation techniques. *Iraqi Journal for Electrical and Electronic Engineering*, v. 17, p. 166–175, 12 2021.

[Adams e Bischof 1994]ADAMS, R.; BISCHOF, L. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 16, n. 6, p. 641–647, 1994.

[Agatonovic-Kustrin e Beresford 2000]AGATONOVIC-KUSTRIN, S.; BERESFORD, R. Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, v. 22, n. 5, p. 717–727, 2000. ISSN 0731-7085. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0731708599002721>.

[Agricultura 2014]AGRICULTURA, P. e. A. MAPA-Ministério da. *Manual de métodos analíticos oficiais para fertilizantes e corretivos*. [S.l.]: MAPA Brasília, 2014.

[5]AL-THYABAT, S.; MILES, N.; KOH, T. Estimation of the size distribution of particles moving on a conveyor belt. *Minerals Engineering*, Elsevier, v. 20, n. 1, p. 72–83, 2007.

[6]ALIPPI, C.; BORACCHI, G.; ROVERI, M. Detecting drops on lens in wireless multimedia sensor network nodes. In: *2009 IEEE International Workshop on Robotic and Sensors Environments*. [S.l.: s.n.], 2009. p. 128 – 133.

[Amit et al. 2022]AMIT, T. et al. *SegDiff: Image Segmentation with Diffusion Probabilistic Models*. 2022. Disponível em: <https://arxiv.org/abs/2112.00390>.

[8]BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 39, n. 12, p. 2481–2495, 2017.

[9]CANFIELD, D. E.; GLAZER, A. N.; FALKOWSKI, P. G. The evolution and future of earth's nitrogen cycle. *science*, American Association for the Advancement of Science, v. 330, n. 6001, p. 192–196, 2010.

[Canny 1986]CANNY, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8, n. 6, p. 679–698, 1986.

[Canny 1986]CANNY, J. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 6, p. 679–698, 1986.

[Cavicchioli et al. 2019]CAVICCHIOLI, R. et al. Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology*, Nature Publishing Group, v. 17, n. 9, p. 569–586, 2019.

[Cezar et al. 2020]CEZAR, O. M. J. et al. Simulated dataset to verify the overlapping and segregation problem on computer vision granullometry of fertilizers. *IFAC-PapersOnLine*, Elsevier BV, v. 53, n. 2, p. 12032–12037, 2020.

[14]CHEEVASUVIT, F.; MAITRE, H.; VIDAL-MADJAR, D. A robust method for picture segmentation based on a split-and-merge procedure. *Computer Vision, Graphics, and Image Processing*, v. 34, n. 3, p. 268–281, 1986. ISSN 0734-189X. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0734189X86800421>.

[Cheng et al. 2022]CHENG, B. et al. Masked-attention mask transformer for universal image segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2022. p. 1290–1299.

[Ferreira e Rasband 2012]FERREIRA, T.; RASBAND, W. Imagej user guide. *ImageJ/Fiji*, v. 1, 2012.

[Foster e Custodio 2019]FOSTER, S.; CUSTODIO, E. Groundwater resources and intensive agriculture in europe–can regulatory agencies cope with the threat to sustainability? *Water Resources Management*, Springer, v. 33, n. 6, p. 2139–2151, 2019.

[Fuglie et al. 2019]FUGLIE, K. et al. *Harvesting prosperity: Technology and productivity growth in agriculture*. [S.l.]: World Bank Publications, 2019.

[Ghosh et al. 2019]GHOSH, S. et al. *Understanding Deep Learning Techniques for Image Segmentation*. arXiv, 2019. Disponível em: <https://arxiv.org/abs/1907.06119>.

[Goulart et al. 2020]GOULART, D. A. et al. Grain surface simulator to averiguate the overlapping and noise problems on computer vision granullometry of fertilizers. In: IEEE. *18th International Conference on Industrial Informatics (INDIN)*. [S.l.], 2020. v. 1, p. 374–379.

[He et al. 2016]HE, K. et al. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778.

[Hu et al. 2020]HU, J. et al. Design and study on lightweight organic fertilizer distributor. *Computers and Electronics in Agriculture*, Elsevier, v. 169, p. 105149, 2020.

[Huang et al. 2017]HUANG, G. et al. Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017.

[Huang et al. 2019]HUANG, Y. et al. Segmentation of overlapping cervical smear cells based on u-net and improved level set. In: *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. [S.l.: s.n.], 2019. p. 3031–3035.

[Igathinathane et al. 2008]IGATHINATHANE, C. et al. Shape identification and particles size distribution from basic shape parameters using imagej. *Computers and Electronics in Agriculture*, v. 63, n. 2, p. 168–182, 2008. ISSN 0168-1699. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0168169908000793>.

[Igathinathane et al. 2009]IGATHINATHANE, C. et al. Sieveless particle size distribution analysis of particulate materials through computer vision. *Computers and Electronics in Agriculture*, Elsevier, v. 66, n. 2, p. 147–158, 2009.

[Jain et al. 2022]JAIN, J. et al. *OneFormer: One Transformer to Rule Universal Image Segmentation*. 2022. Disponível em: <https://arxiv.org/abs/2211.06220>.

[Kattenborn et al. 2021]KATTENBORN, T. et al. Review on convolutional neural networks (cnn) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 173, p. 24–49, 2021. ISSN 0924-2716. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0924271620303488>.

[Kirillov et al. 2023]KIRILLOV, A. et al. Segment anything. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2023. p. 4015–4026.

[Kondermann 2013]KONDERMANN, D. Ground truth design principles: An overview. In: . [S.l.: s.n.], 2013.

[Kramer 1991]KRAMER, M. A. Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, v. 37, n. 2, p. 233–243, 1991. Disponível em: <https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.690370209>.

[32]KURT, B.; NABIYEV, V. V.; TURHAN, K. Medical images enhancement by using anisotropic filter and clahe. In: IEEE. *Innovations in Intelligent Systems and Applications (INISTA), 2012 International Symposium On*. [S.l.], 2012. p. 1–4.

[Le et al. 2018]LE, T.-T. et al. Morphology characterization of irregular particles using image analysis. application to solid inorganic fertilizers. *Computers and Electronics in Agriculture*, v. 147, p. 146–157, 2018. ISSN 0168-1699. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0168169916308791>.

[Liu e Yu 2009]LIU, D.; YU, J. Otsu method and k-means. In: IEEE. *9th International Conference on Hybrid Intelligent Systems (ICHIS)*. [S.l.], 2009. p. 344–349.

[Lloyd 1982]LLOYD, S. Least squares quantization in pcm. *IEEE transactions on information theory*, IEEE, v. 28, n. 2, p. 129–137, 1982.

[Ma et al. 2020]MA, X. et al. A new belt ore image segmentation method based on the convolutional neural network and the image-processing technology. *Minerals*, v. 10, n. 12, 2020. ISSN 2075-163X. Disponível em: <https://www.mdpi.com/2075-163X/10/12/1115>.

[Maiti et al. 2017]MAITI, A. et al. Development of a mass model in estimating weight-wise particle size distribution using digital image processing. *International Journal of Mining Science and Technology*, Elsevier, v. 27, n. 3, p. 435–443, 2017.

[Maja e Ayano 2021]MAJA, M. M.; AYANO, S. F. The impact of population growth on natural resources and farmers' capacity to adapt to climate change in low-income countries. *Earth Systems and Environment*, Springer, v. 5, n. 2, p. 271–283, 2021.

[Mavridou et al. 2019]MAVRIDOU, E. et al. Machine vision systems in precision agriculture for crop farming. *Journal of Imaging*, Multidisciplinary Digital Publishing Institute, v. 5, n. 12, p. 89, 2019.

[Melo et al. 2014]MELO, G. J. A. de et al. Método de limiarização automática para a contagem de células em imagens microscópicas. *Revista GEINTEC-Gestão, Inovação e Tecnologias*, v. 4, n. 3, p. 1283–1291, 2014.

[Meng et al. 2018]MENG, Y. et al. Automatic detection of particle size distribution by image analysis based on local adaptive canny edge detection and modified circular hough transform. *Micron*, Elsevier, v. 106, p. 34–41, 2018.

[Microtrac 2025]MICROTRAC. *Camsizer Online: Imagem do CAMSIZER para NPK e outros granulares*. 2025. https://www.microtrac.com/products/particle-size-shape-analysis/dynamic-image-analysis/camsizer-online/. [imagem digital]. Acesso em: 05 mar. 2025.

[Norozpour e Momenzadeh 2021]NOROZPOUR, S.; MOMENZADEH, M. On comparison of different image segmentation techniques. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, v. 12, p. 4659–4663, 01 2021.

[Otsu 1979]OTSU, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, IEEE, v. 9, n. 1, p. 62–66, 1979.

[Rahman et al. 2016]RAHMAN, S. et al. An adaptive gamma correction for image enhancement. *EURASIP Journal on Image and Video Processing*, v. 2016, n. 1, p. 35, 2016. ISSN 1687-5281. Disponível em: <https://doi.org/10.1186/s13640-016-0138-1>.

[Rahmawati et al. 2021]RAHMAWATI, S. et al. Prewitt and canny methods on inversion image edge detection: An evaluation. *Journal of Physics: Conference Series*, IOP Publishing, v. 1933, n. 1, p. 012039, jun 2021. Disponível em: <https://dx.doi.org/10.1088/1742-6596/1933/1/012039>.

[47]RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. [S.l.: s.n.], 2015. p. 234–241.

[48]RUIJVEN, B. J. V.; CIAN, E. D.; WING, I. S. Amplification of future energy demand growth due to climate change. *Nature Communications*, Nature Publishing Group, v. 10, n. 1, p. 1–12, 2019.

[Sarkar et al. 2018]SARKAR, P. et al. Pore system, microstructure and porosity characterization of gondwana shale of eastern india using laboratory experiment and watershed image segmentation algorithm. *Marine and Petroleum Geology*, v. 94, p. 246–260, 2018. ISSN 0264-8172. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0264817218301582>.

[Sharma et al. 2020]SHARMA, R. et al. A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Computers & Operations Research*, Elsevier, v. 119, p. 104926, 2020.

[51]SHARMA, S.; SHARMA, S.; ATHAIYA, A. Activation functions in neural networks. *towards data science*, v. 6, n. 12, p. 310–316, 2017.

[Shorten e Khoshgoftaar 2019]SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. *J. Big Data*, v. 6, p. 60, 2019. Disponível em: <https://doi.org/10.1186/s40537-019-0197-0>.

[Simonyan e Zisserman 2014]SIMONYAN, K.; ZISSERMAN, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv, 2014. Disponível em: <https://arxiv.org/abs/1409.1556>.

[Smith et al. 2013]SMITH, P. et al. How much land-based greenhouse gas mitigation can be achieved without compromising food security and environmental goals? *Global Change Biology*, Wiley Online Library, v. 19, n. 8, p. 2285–2302, 2013.

[Steffen et al. 2015]STEFFEN, W. et al. Planetary boundaries: Guiding human development on a changing planet. *Science*, American Association for the Advancement of Science, v. 347, n. 6223, p. 1259855, 2015.

[56]TAFESSE, S.; FERNLUND, J.; BERGHOLM, F. Digital sieving-matlab based 3-d image analysis. *Engineering Geology*, Elsevier, v. 137, p. 74–84, 2012.

[Tehreem et al. 2020]TEHREEM, H. S. et al. Impact of average temperature, energy demand, sectoral value added, and population growth on water resource quality and mortality rate: it is time to stop waiting around. *Environmental Science and Pollution Research*, Springer, v. 27, n. 30, p. 37626–37644, 2020.

[Tobin et al. 2017]TOBIN, J. et al. *Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World*. 2017. Disponível em: <https://arxiv.org/abs/1703.06907>.

[Traversi et al. 2024]TRAVERSI, N. D. F. et al. Development of comprehensive fertilizer datasets: Enhancing precision agriculture through data-driven insights. In: *2024 IEEE 22nd International Conference on Industrial Informatics (INDIN)*. [S.l.]: IEEE, 2024. p. 1–6.

[Traversi et al. 2022]TRAVERSI, N. d. F. et al. A neural network for segmentation of fertilizer grain with multiple sizes and without background. In: *IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society*. [S.l.]: IEEE, 2022. p. 1–6.

[Traversi et al. 2021]TRAVERSI, N. d. F. et al. VBM granulometry for real applications in the fertilizer industry. In: *2021 Latin American Robotics Symposium (LARS), 2021 Brazilian Symposium on Robotics (SBR), and 2021 Workshop on Robotics in Education (WRE)*. [S.l.]: IEEE, 2021.

[Traversi et al. 2021]TRAVERSI, N. d. F. et al. Vbm granulometry for real applications in the fertilizer industry. In: IEEE. *18th Latin American Robotics Symposium (LARS)*. [S.l.], 2021. p. 276–281.

[Tyagi 2018]TYAGI, V. *Understanding digital image processing*. [S.l.]: CRC Press, 2018.

[64]VINCENT, O. R.; FOLORUNSO, O. et al. A descriptive algorithm for sobel image edge detection. In: *Proceedings of informing science & IT education conference (InSITE)*. [S.l.: s.n.], 2009. v. 40, p. 97–107.

[Vollset et al. 2020]VOLLSET, S. E. et al. Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for

the global burden of disease study. *The Lancet*, Elsevier, v. 396, n. 10258, p. 1285–1306, 2020.

[Wang et al. 2021]WANG, J. et al. Novel water-and fertilizer-management strategy: Nutrient-water carrier. *Journal of Cleaner Production*, Elsevier, v. 291, p. 125961, 2021.

[67]WEISS, K.; KHOSHGOFTAAR, T. M.; WANG, D. A survey of transfer learning. *Journal of Big Data*, v. 3, n. 1, p. 9, May 2016. ISSN 2196-1115. Disponível em: <https://doi.org/10.1186/s40537-016-0043-6>.

[Wen et al. 2021]WEN, Q. et al. Time series data augmentation for deep learning: A survey. In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, 2021. Disponível em: <https://doi.org/10.24963

[Wu e Li 2022]WU, Y.; LI, Q. The algorithm of watershed color image segmentation based on morphological gradient. *Sensors*, MDPI AG, v. 22, n. 21, p. 8202, out. 2022. Disponível em: <https://doi.org/10.3390/s22218202>.

[Wu et al. 2017]WU, Y. et al. Improved image segmentation method based on morphological reconstruction. *Multimedia Tools and Applications*, v. 76, n. 19, p. 19781–19793, Oct 2017. ISSN 1573-7721. Disponível em: <https://doi.org/10.1007/s11042-015-3192-2>.

[Xie et al. 2021]XIE, E. et al. *SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers*. 2021. Disponível em: <https://arxiv.org/abs/2105.15203>.

[72]YANG, K.; ZHONG, W.; LI, F. Leaf segmentation and classification with a complicated background using deep learning. *Agronomy*, Multidisciplinary Digital Publishing Institute, v. 10, n. 11, p. 1721, 2020.

[Zeng 2024]ZENG, W. Image data augmentation techniques based on deep learning: A survey. *Mathematical Biosciences and Engineering*, 2024.

[Zhan e Zhang 2019]ZHAN, Y.; ZHANG, G. An improved otsu algorithm using histogram accumulation moment for ore segmentation. *Symmetry*, v. 11, n. 3, 2019. ISSN 2073-8994. Disponível em: <https://www.mdpi.com/2073-8994/11/3/431>.