

Detecção de Doenças em Folhas com Reconstrução por Pix2Pix, Métrica CIEDE2000 e Interpretação via Grad-CAM

Pedro Marcinoni Varão Ribeiro Vieira da Rosa
Universidade de Brasília (UnB)
Disciplina: Introdução à Inteligência Artificial
Email: pedroextrer@gmail.com

Leonardo Mileo Krauss
Universidade de Brasília (UnB)
Disciplina: Introdução à Inteligência Artificial
Email: leomileoguima@gmail.com

Resumo—Este trabalho investiga a detecção não supervisória de doenças em folhas utilizando reconstrução via Pix2Pix treinado exclusivamente com imagens saudáveis. A hipótese central é que folhas doentes apresentam padrões cromáticos e estruturais que dificultam a reconstrução pelo modelo, de modo que a diferença entre imagem real e reconstruída, medida pela métrica CIEDE2000, pode atuar como indicador robusto de anomalia. Além disso, incorporamos Grad-CAM para inspecionar como o discriminador PatchGAN avalia regiões relevantes, permitindo interpretação visual dos padrões considerados anômalos. Os resultados alcançados demonstram desempenho competitivo para um dataset com imperfeições estruturais, obtendo AUC-ROC de 0,901 e F1-score de 0,8768.

I. INTRODUÇÃO

A detecção automática de doenças em plantas é um problema central na agricultura de precisão. Métodos supervisionados dependem de grandes conjuntos anotados, que frequentemente não estão disponíveis e apresentam elevada variabilidade entre espécies, iluminação e condições ambientais. Alternativamente, modelos generativos adversariais (GANs) têm sido aplicados para detecção não supervisionada de anomalias, reconstruindo apenas amostras saudáveis e utilizando a discrepância da reconstrução como indicador de doença.

Neste trabalho, seguimos a abordagem baseada em Pix2Pix [2], tal como apresentada no artigo base disponibilizado em aula [1]. O gerador é treinado para colorizar folhas saudáveis a partir de imagens em escala de cinza, produzindo uma reconstrução que, quando comparada à imagem real via CI-EDE2000, fornece um *anomaly score*. O limiar utilizado é definido pela maximização do F1-score.

Também incorporamos Grad-CAM para interpretação visual no discriminador, a fim de avaliar quais regiões da folha influenciam mais a decisão adversarial.

II. TRABALHOS RELACIONADOS

Este trabalho baseia-se no repositório oficial `pytorch-CycleGAN-and-pix2pix` [3], amplamente utilizado em tarefas de tradução de imagens e reconstrução. A metodologia segue o artigo fornecido pelo professor [1], que emprega reconstrução adversarial para detecção de anomalias em folhas.

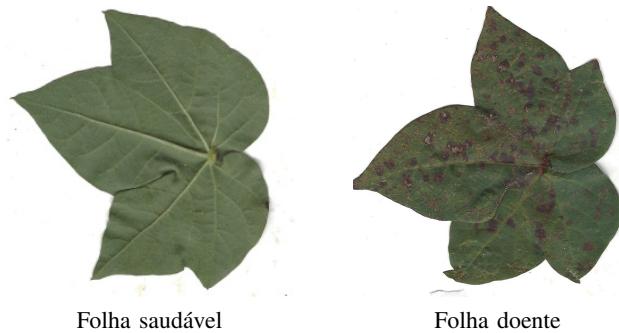
III. DATASET E PRÉ-PROCESSAMENTO

O dataset fornecido contém 100 imagens saudáveis e 100 doentes, apresentando variação de proporção, fundo não uniforme e, em alguns casos, objetos próximos às folhas.

Para o treinamento, foram selecionadas aleatoriamente 50 folhas saudáveis. O conjunto de teste foi composto pelas 50 folhas saudáveis restantes e pelas 100 folhas doentes.

As imagens foram convertidas automaticamente para o espaço LAB pelo código-base, e um script adicional foi desenvolvido para redimensioná-las para 256×256 pixels. A normalização seguiu o padrão $[-1, 1]$ utilizado pelo repositório.

Figura 1. Exemplo de imagens do dataset: folha saudável (esquerda) e folha doente (direita).



IV. MODELO E CONFIGURAÇÃO DE TREINO

O modelo utilizado foi o Pix2Pix padrão, composto por:

- gerador U-Net 256 (netG=unet_256);
 - discriminador PatchGAN 70x70 (netD=basic).

A perda total foi definida como:

$$\mathcal{L} = \mathcal{L}_{GAN} + \lambda \mathcal{L}_{I,J} \quad \lambda = 5$$

O otimizador adotado foi Adam com taxa de aprendizado de 0,00015, $\beta_1 = 0,5$ e batch size igual a 1. Embora o plano inicial fosse treinar por 150 épocas (100 estáveis + 50 de decaimento linear), verificou-se que após a época 15 o modelo começava a produzir cores irreais, reduzindo acentuadamente o

AUC. Esse comportamento decorre do fato de que, em estágios avançados do treino, o gerador passa a enfatizar padrões cromáticos artificiais para maximizar a perda adversarial, produzindo colorações que “enganam” o discriminador mas se distanciam do espaço cromático fisiológico aprendido a partir das folhas saudáveis. Utilizou-se, portanto, early stopping na época 15.

V. ESTRUTURA DO REPOSITÓRIO

O repositório segue uma organização inspirada no projeto oficial `pytorch-CycleGAN-and-pix2pix`, com separação explícita entre scripts, dados e resultados. Os elementos principais são:

- `train.py`: treinamento completo do modelo Pix2Pix.
- `scripts/test_model.py`: inferência em todo o conjunto de teste, salvando reconstruções, mapas CIEDE2000 e métricas.
- `scripts/test_single_leaf.py`: inferência em uma única imagem.
- `scripts/show_GradCAM.py`: aplicação de Grad-CAM no discriminador.
- `datasets/`: imagens pré-processadas.
- `source_images/`: imagens originais não processadas.
- `checkpoints/`: pesos salvos e logs de treino.
- `results/`: reconstruções, mapas CIEDE2000, Grad-CAM e métricas.

Essa estrutura garante modularidade e reproduzibilidade dos experimentos.

VI. REPRODUTIBILIDADE E EXECUÇÃO

Todos os experimentos podem ser reproduzidos utilizando os scripts descritos nesta seção. As dependências principais constam no arquivo `environment.yml`.

A. Treinamento

Figura 2. Comando utilizado para o treinamento do modelo.

```
python train.py
--dataroot
D:/pytorch-CycleGAN-and-pix2pix/
datasets/leaf_disease_detection
--name pix2pix_final_v3
--model colorization
--dataset_mode colorization
--direction AtoB
--lr 0.00015
--lambda_L1 5.0
--beta1 0.5
--n_epochs 100
--n_epochs_decay 50
--netG unet_256
--netD basic
--num_threads 0
```

B. Teste em todo o conjunto

Figura 3. Comando de inferência para todo o conjunto de teste.

```
python scripts/test_model.py
--dataroot
./datasets/leaf_disease_detection
```

C. Teste em uma única imagem

Figura 4. Comando de inferência para uma única imagem.

```
python scripts/test_single_leaf.py
--dataroot
./datasets/leaf_disease_detection
--path
"../datasets/leaf_disease_detection/
test/doentes/a988-992_ab_0.jpg"
```

D. Gerar visualizações Grad-CAM

Figura 5. Comando utilizado para gerar visualizações Grad-CAM.

```
python scripts/show_GradCAM.py
--dataroot
./datasets/leaf_disease_detection
```

VII. PROCEDIMENTO DE INFERÊNCIA E CÁLCULO DO SCORE

O processo de inferência tem como objetivo quantificar o quanto o modelo, treinado exclusivamente com amostras saudáveis, é capaz de reconstruir adequadamente uma nova imagem. Parte-se da hipótese de que regiões doentes, por exibirem padrões cromáticos e estruturais que não pertencem à distribuição aprendida, gerarão maiores discrepâncias entre a imagem original e a reconstruída. Tais discrepâncias são mensuradas pela métrica CIEDE2000, cujo acúmulo define o *anomaly score* final.

O script `test_model.py` implementa integralmente o pipeline de avaliação, composto pelas seguintes etapas:

- 1) a imagem RGB é carregada e normalizada;
- 2) aplica-se a transformação padronizada para o espaço de cor LAB, separando-se o canal de luminância L dos canais cromáticos a, b ;
- 3) o gerador recebe apenas o canal L e estima os canais cromáticos \hat{a}, \hat{b} ;
- 4) a reconstrução LAB (L, \hat{a}, \hat{b}) é convertida novamente para o espaço RGB;
- 5) calcula-se, para cada pixel i , a diferença cromática

$$\Delta E_i = \Delta E_{CIEDE2000}((L_i, a_i, b_i), (L_i, \hat{a}_i, \hat{b}_i));$$

- 6) o *anomaly score* global é obtido pela agregação dos erros locais,

$$S = \sum_i \Delta E_i,$$

refletindo o grau de incompatibilidade entre a imagem real e sua reconstrução.

O valor resultante é então comparado a um limiar previamente determinado pela maximização do F1-score, classificando a imagem como saudável ou doente.

A opção pelo F1-score, em vez do índice de Youden, decorre do desbalanceamento do conjunto de dados e da necessidade de priorizar a detecção de folhas doentes. O índice de Youden busca equilibrar sensibilidade e especificidade, mas no nosso caso produzia um limiar com recall reduzido (0,7600), aumentando falsos negativos. Como a tarefa exige identificar anomalias de forma robusta, o F1-score, que penaliza desbalanceamentos entre precisão e recall, forneceu um limiar mais adequado ao objetivo prático do modelo.

VIII. RESULTADOS

As métricas obtidas para as 150 imagens avaliadas (50 saudáveis e 100 doentes) são apresentadas na Tabela I.

Tabela I
MÉTRICAS DE DESEMPENHO DO MODELO.

Métrica	Valor
AUC-ROC	0.9010
Acurácia	0.8333
Precisão	0.8641
Recall	0.8900
F1-score	0.8768
Limiar F1-max	1.7391

A matriz de confusão correspondente é:

	Pred. Neg.	Pred. Pos.
Real Neg.	36	14
Real Pos.	11	89

A. Avaliação Quantitativa dos Resultados

As métricas obtidas evidenciam que o modelo apresenta desempenho consistente na distinção entre folhas saudáveis e doentes. O AUC-ROC de 0,9010 indica boa separabilidade global entre as classes ao longo de todos os limiares possíveis. A acurácia (0,8333), precisão (0,8641) e recall (0,8900) mostram que o sistema mantém equilíbrio entre detecção correta de anomalias e controle de falsos positivos. O F1-score de 0,8768 sintetiza esse equilíbrio, refletindo robustez mesmo em um conjunto de teste com leves irregularidades. O limiar ótimo encontrado (1,7391) maximiza esse compromisso entre sensibilidade e especificidade.

B. Análise Qualitativa dos resultados

A análise visual mostrou que, em folhas saudáveis, o modelo preservou a textura natural e manteve boa estabilidade cromática. Em folhas doentes, ocorreram deformações cromáticas e artefatos.



Figura 6. Da esquerda para a direita: imagem em BW, imagem original, reconstruída e mapa CIEDE2000 de uma folha doente

As regiões de maior intensidade no mapa CIEDE2000 evidenciam discrepâncias locais entre a imagem original e a reconstruída pelo gerador. Como o Pix2Pix foi treinado exclusivamente com amostras saudáveis, sua reconstrução tende a projetar cores e texturas compatíveis com o padrão fisiológico aprendido. Dessa forma, estruturas patológicas não são fielmente reproduzidas, resultando em erros elevados nesses pontos. As regiões quentes, portanto, correspondem a áreas onde o modelo apresenta maior dificuldade de reconstrução, revelando de maneira explícita os indícios de anomalia. Além disso, é importante notar que a etiqueta da imagem recebeu um alto score de discrepança, o que não influencia no caso dessa classificação correta, mas pode gerar falsos positivos com mais frequência.



Figura 7. Da esquerda para a direita: imagem em BW, imagem original, reconstruída e mapa CIEDE2000 de uma folha saudável

Nas imagens de folhas saudáveis, observa-se que o mapa CIEDE2000 apresenta predominantemente regiões frias, indicando baixa discrepança entre a imagem original e a reconstruída. Como o Pix2Pix foi treinado apenas com amostras não patológicas, o gerador reproduz com elevada fidelidade as cores e texturas características do padrão saudável, resultando em erros mínimos ao comparar os espaços cromáticos. As poucas regiões levemente aquecidas tendem a corresponder a variações naturais de iluminação, reflexos ou irregularidades sutis da superfície foliar, e não a anomalias estruturais. Assim, a homogeneidade do mapa reforça a consistência do modelo na reconstrução de casos normais e evidencia a separabilidade entre padrões saudáveis e doentes.

IX. INTERPRETAÇÃO VIA GRAD-CAM

Aplicamos Grad-CAM às cinco camadas convolucionais intermediárias do discriminador PatchGAN, com o objetivo de identificar quais regiões espaciais mais contribuem para o julgamento adversarial. Diferentemente do gerador, cujo papel é reconstruir a cromaticidade, o discriminador opera localmente em janelas de 70×70 pixels, avaliando padrões texturais e cromáticos de alta e média frequência. Assim, suas ativações revelam como a arquitetura internaliza características que distinguem imagens reais das geradas.

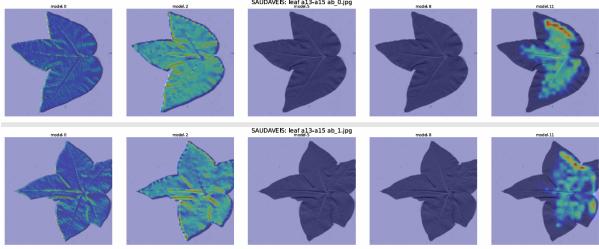


Figura 8. Exemplo de Grad-CAM aplicado ao discriminador em duas folhas saudáveis.

As visualizações revelaram:

- **Primeira camada:** observa-se um mapa predominantemente frio, porém seguindo claramente o contorno da folha. Isso indica que, nas primeiras convoluções, o discriminador responde majoritariamente à presença geométrica do objeto, mas com baixa intensidade de ativação, refletindo uma leitura ainda superficial de bordas e contraste global.
- **Segunda camada:** surgem regiões levemente mais quentes, especialmente em áreas mais escuras da superfície. A rede passa a enfatizar variações locais de textura e shading, sugerindo que irregularidades tonais, mesmo em folhas saudáveis, já são tratadas como elementos informativos para a distinção adversarial.
- **Terceira e quarta camadas:** praticamente não exibem ativações significativas em ambas as folhas, apresentando mapas de calor quase neutros. Esse fenômeno indica que, para entradas saudáveis, o discriminador não encontra padrões de alta complexidade capazes de gerar respostas fortes nessas profundidades intermediárias. Isso é consistente com a hipótese de que estruturas patológicas, são as responsáveis por fortes ativações nessas camadas quando presentes.
- **Quinta camada:** reaparece uma ativação localizada, caracterizada por um rastro quente na região superior da folha e uma região em seu interior. Esse tipo de resposta sugere que as últimas camadas se especializam em detectar padrões texturais mais abstratos, e que pequenas flutuações estruturais ou de iluminação podem gerar respostas pontuais mesmo em folhas saudáveis. A recorrência dessa ativação na segunda folha reforça que o discriminador é sensível a variações finas do relevo superficial e do gradiente de luminosidade.

X. DISCUSSÃO

Os resultados obtidos demonstram que a abordagem baseada em reconstrução adversarial combinada com a métrica perceptual $\Delta E_{CIEDE2000}$ foi capaz de diferenciar de maneira consistente folhas saudáveis e doentes, mesmo sem supervisão explícita. O modelo alcançou AUC-ROC de 0,901, F1-score de 0,8768 e precisão de 0,8641, valores que indicam um desempenho robusto dentro do contexto experimental. Observa-se particularmente que o recall atingiu 0,8900, sugerindo que o método é eficaz em identificar imagens doentes, objetivo central de um sistema de detecção de anomalias.

A matriz de confusão evidencia que a maioria dos erros ocorreu na classificação de folhas saudáveis, que em alguns casos receberam pontuações de anomalia elevadas. A análise qualitativa indica que esses falsos positivos estão associados principalmente à presença de fundo heterogêneo, sombras e etiquetas. Esse comportamento pode ser observado na Figura 6, quando a etiqueta é marcada com alto score de discrepância

Ainda assim, o modelo apresentou forte coerência interna: folhas saudáveis foram reconstruídas de forma estável e homogênea, enquanto folhas doentes exibiram altas discrepâncias localizadas, refletidas em mapas CIEDE2000 intensos nas áreas correspondentes às lesões. Tais evidências sugerem que, mesmo sem segmentação e padronização de dataset, como no artigo original, o modelo aprendeu representações estruturais significativas do padrão cromático normal.

XI. CONCLUSÃO

Este trabalho demonstrou que a combinação de reconstrução via Pix2Pix com a métrica perceptual CIEDE2000 constitui uma ferramenta eficaz para detecção não supervisionada de doenças em folhas. O modelo alcançou desempenho competitivo e apresentou comportamento consistente tanto nas métricas quantitativas quanto nas análises qualitativas. As visualizações por Grad-CAM revelaram padrões interpretáveis de atenção no discriminador, contribuindo para a validação da abordagem sob uma perspectiva explicável.

Embora alguns falsos positivos tenham sido induzidos por elementos externos como etiquetas e irregularidades no fundo, esses fatores não comprometeram a capacidade geral do método de distinguir padrões fisiológicos normais de anomalias. Melhorias futuras podem concentrar-se na uniformização das condições de captura (especialmente remoção de etiquetas e padronização do fundo), mas os resultados indicam que a metodologia é robusta mesmo sem pré-processamento intenso.

REFERÊNCIAS

- [1] R. Katafuchi and T. Tokunaga, “Image-based Plant Disease Diagnosis with Unsupervised Anomaly Detection Based on Reconstructability of Colors,” arXiv preprint arXiv:2011.14306, 2020. Disponível em: <https://arxiv.org/pdf/2011.14306.pdf>.
- [2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *CVPR*, 2017.
- [3] J.-Y. Zhu et al., “pytorch-CycleGAN-and-pix2pix,” GitHub repository. Disponível em: <https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix>.