

Classificação de Objetos com CNNs e Autoencoder

em Dataset Derivado de Detecção de Focos de Aedes Aegypti

CPE 727 - Deep Learning

Felipe Barreto Andrade (felipebarretoandrade@poli.ufrj.br)

December 9, 2025

Agenda

o

- ▶ Introdução
- ▶ Problema e Base de Dados
- ▶ Método Proposto
- ▶ Resultados Obtidos
- ▶ Conclusões
- ▶ Referências Bibliográficas

Agenda

1 Introdução

- ▶ Introdução
- ▶ Problema e Base de Dados
- ▶ Método Proposto
- ▶ Resultados Obtidos
- ▶ Conclusões
- ▶ Referências Bibliográficas

Motivação

1 Introdução

- Focos de Aedes aegypti em ambientes urbanos são difíceis de monitorar em larga escala.
- Já existe um dataset de detecção (MBGv2) [1, 2], com anotações de caixas para potenciais criadouros.
- Neste trabalho, derivamos um problema de **classificação de objetos em crops**:
 - Identificar o tipo de objeto (tanque, pneu, poça, piscina, etc.) a partir de recortes de 128×128 .
- Isso permite estudar, de forma controlada, o impacto de arquiteturas de deep learning e regularização.

Objetivo do Trabalho

1 Introdução

- Construir um pipeline de classificação de objetos em imagens derivadas do MBGv2.
- Comparar três abordagens baseadas em deep learning:
 - **CNN base supervisionada.**
 - **CNN com regularização** (dropout + weight decay + data augmentation forte).
 - **Autoencoder convolucional + classificador MLP.**
- Avaliar o impacto dessas escolhas nas figuras de mérito:
 - Acurácia global no teste.
 - Macro-F1 (importante para classes minoritárias).

Macro-F1 em Problemas Multiclasse

1 Introdução

- Em problemas multiclasse, podemos calcular uma métrica F1 por classe (one-vs-rest):

$$F1_c = \frac{2 \cdot \text{precision}_c \cdot \text{recall}_c}{\text{precision}_c + \text{recall}_c}$$

- Seja C o número total de classes do problema (no caso, $C = 12$).
- A **Macro-F1** é a média aritmética dos F1 de todas as classes [3]:

$$\text{Macro-F1} = \frac{1}{C} \sum_{c=1}^C F1_c$$

- Cada classe contribui igualmente, independentemente do número de exemplos.
- Em bases desbalanceadas (como o MBGv2), a Macro-F1 é importante porque:
 - penaliza modelos que ignoram classes minoritárias;
 - complementa a acurácia, que pode ficar alta mesmo com desempenho ruim em classes raras.

Agenda

2 Problema e Base de Dados

- ▶ Introdução
- ▶ Problema e Base de Dados
- ▶ Método Proposto
- ▶ Resultados Obtidos
- ▶ Conclusões
- ▶ Referências Bibliográficas

Problema de Classificação

2 Problema e Base de Dados

- A partir de vídeos de inspeção urbana, o dataset MBGv2 traz anotações de bounding boxes para candidatos a focos de Aedes.
- Transformamos o problema de detecção em:

Tarefa

Dado um crop 128×128 contendo um objeto anotado, prever sua classe entre 12 categorias.

- Classes incluem, por exemplo:
 - watertank, pool, potted_plant, tire, plastic_bag, etc.

Dataset Derivado do MBGv2 - parte 1

2 Problema e Base de Dados

- Partimos do MBGv2, com anotações de bounding boxes em vídeos urbanos (vídeos 10–37).
- Para cada bounding box anotada:
 - recortamos a região correspondente no frame original;
 - redimensionamos o crop para 128×128 pixels, com interpolação bilinear, mantendo 3 canais (RGB);
 - esse redimensionamento padroniza o tamanho de entrada para as CNNs, ainda que introduza pequena distorção de aspecto em alguns casos.
- Filtragem aplicada:
 - descartar boxes cujo frame não é encontrado ou não pôde ser lido;
 - descartar boxes degeneradas (largura ou altura ≤ 0) ou crops vazios;
 - descartar imagens de saída que não puderam ser salvas.

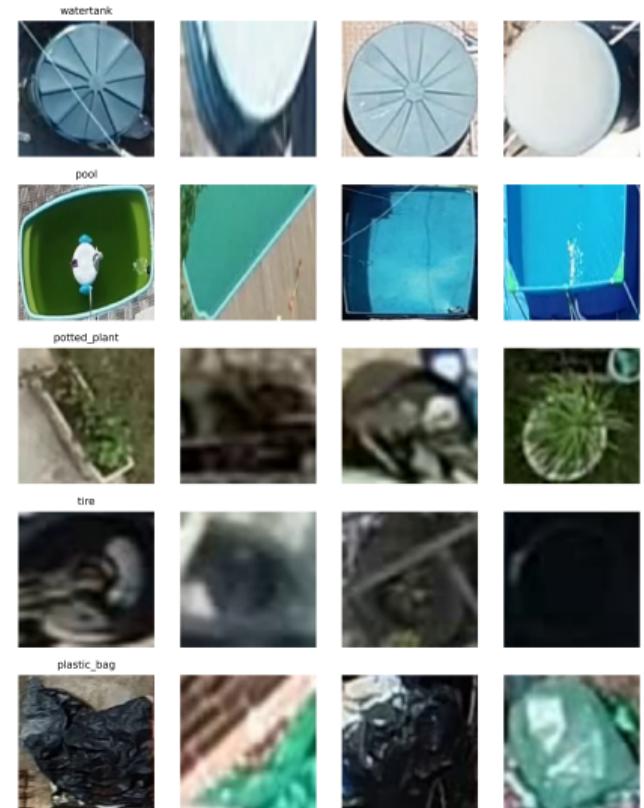
Dataset Derivado do MBGv2 - parte 2

2 Problema e Base de Dados

- Após a filtragem, obtivemos:
 - 57 573 crops de tamanho 128×128 (12 classes).
- Divisão em conjuntos de treino, validação e teste com **split estratificado**:
 - Treino: 40 301 amostras
 - Validação: 8 636 amostras
 - Teste: 8 636 amostras
- Split estratificado preserva, em cada subconjunto, aproximadamente a mesma proporção de amostras de cada classe observada no dataset completo.
- A distribuição de classes é fortemente desbalanceada (por exemplo, watertank muito mais frequente do que dumpster), o que torna relevante o uso de métricas como a **Macro-F1**.

Exemplos de Crops

2 Problema e Base de Dados



- Exemplos por classe ajudam a visualizar a variabilidade intra-classe.
- Ilustra também a presença de ruído de anotação e oclusões.

Agenda

3 Método Proposto

- ▶ Introdução
- ▶ Problema e Base de Dados
- ▶ Método Proposto
- ▶ Resultados Obtidos
- ▶ Conclusões
- ▶ Referências Bibliográficas

Visão Geral do Pipeline

3 Método Proposto

- Etapas principais:
 1. Geração de crops e metadados a partir das anotações do MBGv2.
 2. Divisão treino/val/test estratificada.
 3. Treino da **CNN base** supervisionada.
 4. Treino da **CNN regularizada** (dropout + weight decay + aug).
 5. Treino do **autoencoder convolucional** e extração de embeddings.
 6. Treino de um **MLP** sobre os embeddings do autoencoder.

Modelo 1: CNN Base Supervisionada

3 Método Proposto

- Arquitetura convolucional simples, com blocos:
 - Convoluções 3×3 + ReLU [4] + MaxPooling [5].
 - Camadas totalmente conectadas finais para 12 classes.
- Otimização:
 - Otimizador Adam, taxa de aprendizado fixa (por exemplo, $1e-3$).
 - Treino por 10 épocas, batch size 64.
- Sem data augmentation explícita, apenas normalização/resize.

Modelo 2: CNN com Regularização

3 Método Proposto

- Mesma arquitetura base, mas com:
 - Dropout nas camadas totalmente conectadas.
 - Weight decay no otimizador.
 - Data augmentation mais agressiva:
 - flips horizontais, jitter de brilho/contraste, pequenas rotações, etc.
- Esperado:
 - Mais robusto a variações dos dados.
 - Possível redução de overfitting, com impacto em desempenho médio.

Modelo 3: Autoencoder + Classificador

3 Método Proposto

- **Autoencoder convolucional:**
 - Encoder com 4 blocos conv+pool, reduzindo 128×128 a um vetor latente de dimensão 256.
 - Decoder simétrico com camadas transpostas para reconstruir a imagem.
- Treino não supervisionado:
 - Função de custo: MSE entre imagem de entrada e reconstrução.
- **Classificador MLP:**
 - Recebe o vetor latente (256) como entrada.
 - Duas camadas densas + dropout, saída com 12 classes.

Agenda

4 Resultados Obtidos

- ▶ Introdução
- ▶ Problema e Base de Dados
- ▶ Método Proposto
- ▶ Resultados Obtidos
- ▶ Conclusões
- ▶ Referências Bibliográficas

Figuras de Mérito - Conjunto de Teste

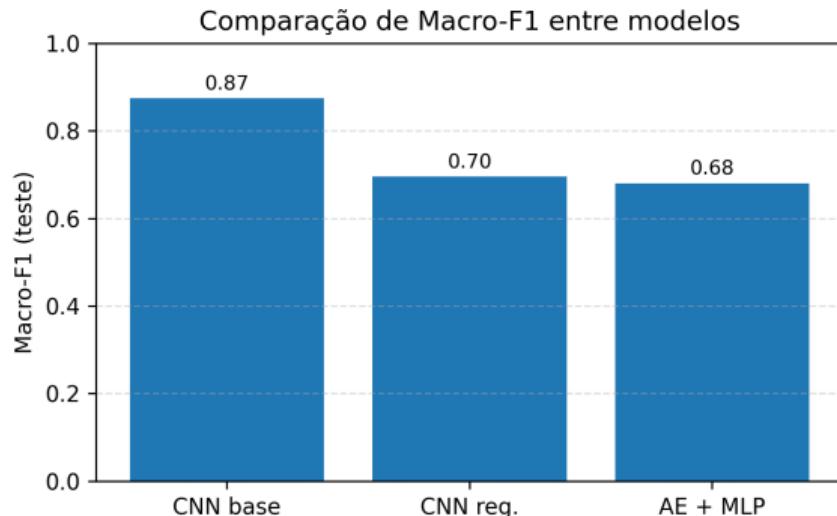
Acurácia e Macro-F1

Modelo	Acurácia	Macro-F1
CNN base (supervisionada)	0.9437	0.8742
CNN com regularização	0.8920	0.6954
Autoencoder + MLP (AE+MLP)	0.8929	0.6802

Table: Resumo das figuras de mérito no conjunto de teste.

Comparação Visual das Figuras de Mérito

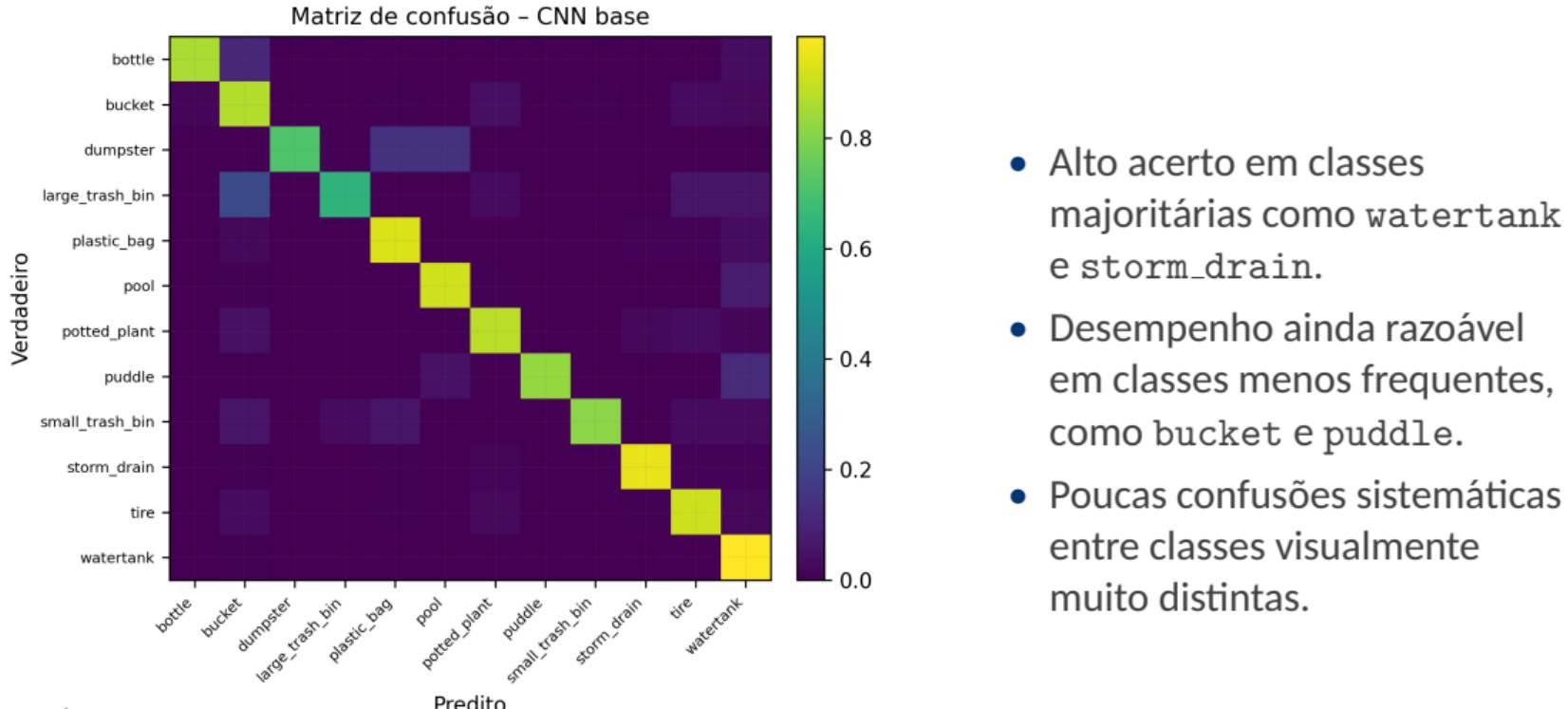
4 Resultados Obtidos



- A CNN base apresenta o melhor desempenho, sobretudo em macro-F1.
- CNN regularizada e AE+MLP têm acurácia semelhante, mas macro-F1 menor.
- Diferenças em macro-F1 refletem o desempenho em classes minoritárias.

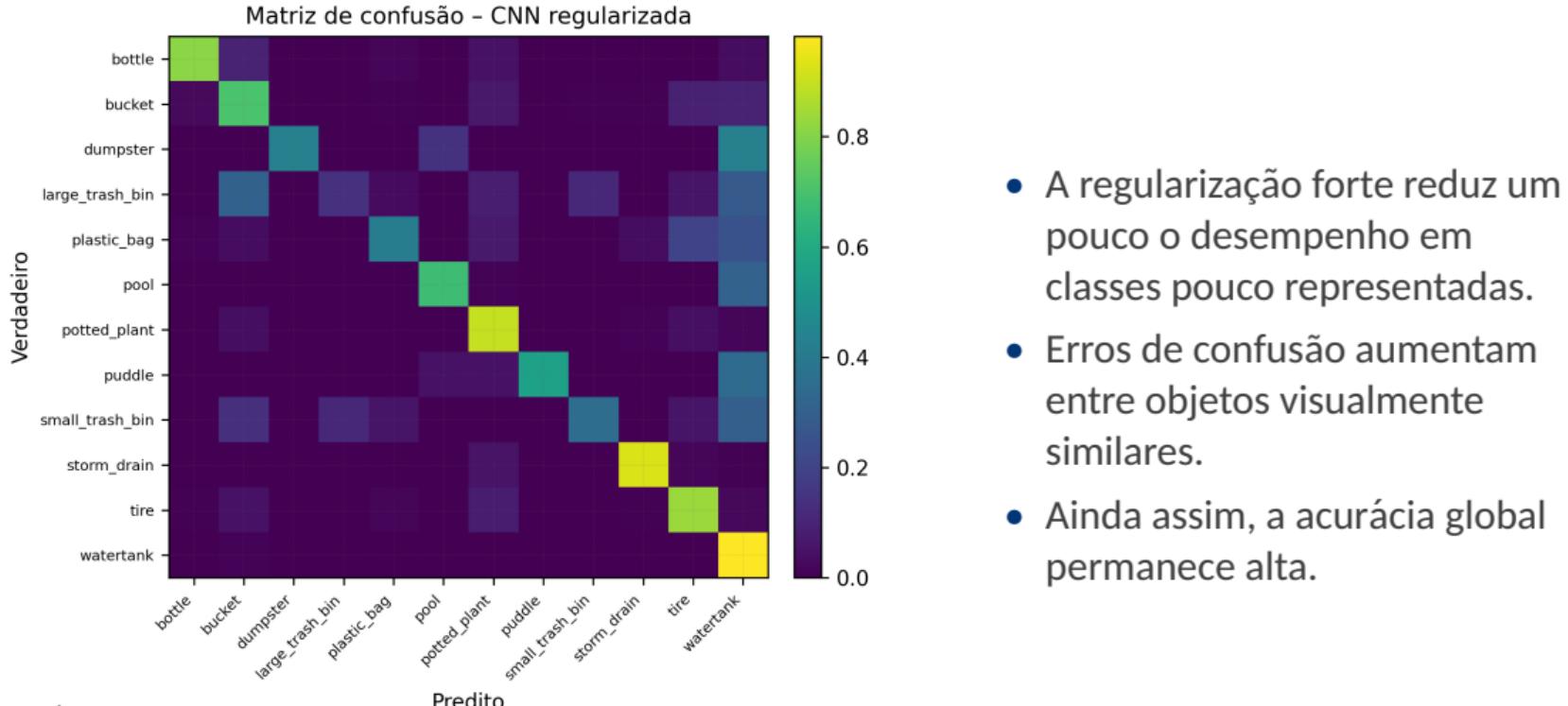
Matriz de Confusão - CNN Base

4 Resultados Obtidos



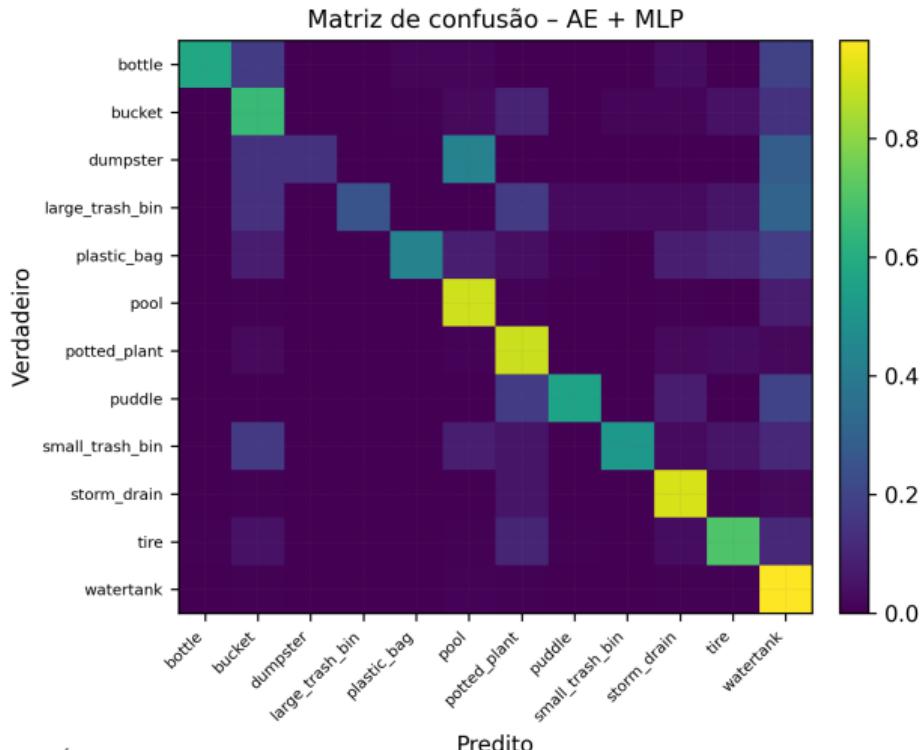
Matriz de Confusão - CNN Regularizada

4 Resultados Obtidos



Matriz de Confusão - Autoencoder + MLP

4 Resultados Obtidos



- Representações aprendidas de forma não supervisionada conseguem suportar bem a classificação.
- Acurácia próxima à CNN regularizada, mas macro-F1 menor.
- Desempenho mais fraco em classes com poucas amostras, como `dumpster` e `large_trash_bin`.

Agenda

5 Conclusões

- ▶ Introdução
- ▶ Problema e Base de Dados
- ▶ Método Proposto
- ▶ Resultados Obtidos
- ▶ Conclusões
- ▶ Referências Bibliográficas

Conclusões

5 Conclusões

- Foi realizado um estudo comparativo de três abordagens de deep learning para classificação de objetos em crops derivados do dataset MBGv2.
- A **CNN base supervisionada** apresentou o melhor desempenho global no conjunto de teste (acurácia 94%, macro-F1 0,87), mostrando que um modelo convolucional relativamente simples, treinado de forma end-to-end, é um baseline forte neste regime de dados.
- A **CNN com regularização** e o **AE+MLP** obtiveram acurácia semelhante, mas macro-F1 menor, indicando que a combinação de regularização mais agressiva e/ou pré-treinamento não supervisionado, da forma utilizada aqui, acabou reduzindo o desempenho principalmente nas classes minoritárias.
- Os resultados ilustram, na prática, que aumentar a complexidade do pipeline (mais regularização, pré-treino, duas etapas) não garante ganhos automáticos, e reforçam a importância de baselines supervisionados bem ajustados.

Trabalhos Futuros

5 Conclusões

- Explorar arquiteturas mais profundas ou pré-treinadas (por exemplo, ResNet-18).
- Balanceamento de classes (reweighting, oversampling, focal loss).
- Integração direta com o problema de detecção (fim-a-fim em bounding boxes).

Agenda

6 Referências Bibliográficas

- ▶ Introdução
- ▶ Problema e Base de Dados
- ▶ Método Proposto
- ▶ Resultados Obtidos
- ▶ Conclusões
- ▶ Referências Bibliográficas

Referências Bibliográficas

6 Referências Bibliográficas

- [1] W. L. Passos, G. M. Araujo, S. L. Netto, and E. A. B. da Silva, “Mbgv2: Mosquito breeding grounds dataset v2.” https://www02.smt.ufrj.br/~tvdigital/database/mosquito/page_02.html, 2020.
Acessado em 09 dez. 2025.
- [2] W. L. Passos, G. M. Araujo, A. A. de Lima, S. L. Netto, and E. A. B. da Silva, “Automatic detection of aedes aegypti breeding grounds based on deep networks with spatio-temporal consistency,” *Computers, Environment and Urban Systems*, vol. 93, p. 101754, 2022.

Referências Bibliográficas

6 Referências Bibliográficas

- [3] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.
- [4] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 2010.
- [5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

Classificação de Objetos com CNNs e Autoencoder

Obrigado pela Atenção!

Alguma Pergunta?

Felipe Barreto Andrade

felipebarretoandrade@poli.ufrj.br