



DISSERTAÇÃO DE MESTRADO

**UMA ARQUITETURA DE *FEW-SHOT LEARNING*  
PARA CLASSIFICAÇÃO DE INSETOS NA AGRICULTURA  
USANDO POUCAS AMOSTRAS**

**JACÓ CIRINO GOMES**

**Brasília, Setembro de 2022**

**UNIVERSIDADE DE BRASÍLIA**

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA  
Faculdade de Tecnologia

DISSERTAÇÃO DE MESTRADO

**UMA ARQUITETURA DE *FEW-SHOT LEARNING*  
PARA CLASSIFICAÇÃO DE INSETOS NA AGRICULTURA  
USANDO POUCAS AMOSTRAS**

**JACÓ CIRINO GOMES**

*Dissertação de Mestrado submetida ao Departamento de Engenharia Mecânica  
como requisito parcial para obtenção do grau de Mestre em Sistemas Mecatrônicos*

Banca Examinadora

Prof. Dr. Díbio Leandro Borges

*Orientador, UnB*

\_\_\_\_\_

Prof. Dr. Fabrizio Alphonsus Alves de Melo Nunes

Soares

*Examinador externo, UFG*

\_\_\_\_\_

Prof. Dr. Flavio Lemes Fernandes

*Examinador externo, UFV*

\_\_\_\_\_

## FICHA CATALOGRÁFICA

CIRINO GOMES, JACÓ

UMA ARQUITETURA DE *FEW-SHOT LEARNING* PARA CLASSIFICAÇÃO DE INSETOS NA AGRICULTURA USANDO POUCAS AMOSTRAS [Distrito Federal] 2022.

xvi, 63 p., 210 x 297 mm (ENM/FT/UnB, Mestre, Engenharia Mecânica, 2022).

Dissertação de Mestrado - Universidade de Brasília, Faculdade de Tecnologia.

Departamento de Engenharia Mecânica

1. Palavra 1

2. Palavra 2

3. Palavra 3

4. Palavra 4

I. ENM/FT/UnB

II. Título (série)

## REFERÊNCIA BIBLIOGRÁFICA

CIRINO GOMES, J. (2022). *UMA ARQUITETURA DE FEW-SHOT LEARNING PARA CLASSIFICAÇÃO DE INSETOS NA AGRICULTURA USANDO POUCAS AMOSTRAS*. Dissertação de Mestrado, Departamento de Engenharia Mecânica, Universidade de Brasília, Brasília, DF, 63 p.

## CESSÃO DE DIREITOS

AUTOR: JACÓ CIRINO GOMES

TÍTULO: UMA ARQUITETURA DE *FEW-SHOT LEARNING* PARA CLASSIFICAÇÃO DE INSETOS NA AGRICULTURA USANDO POUCAS AMOSTRAS .

GRAU: Mestre em Sistemas Mecatrônicos ANO: 2022

É concedida à Universidade de Brasília permissão para reproduzir cópias desta Dissertação de Mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. Os autores reservam outros direitos de publicação e nenhuma parte dessa Dissertação de Mestrado pode ser reproduzida sem autorização por escrito dos autores.

---

JACÓ CIRINO GOMES

Depto. de Engenharia Mecânica (ENM) - FT

Universidade de Brasília (UnB)

Campus Darcy Ribeiro

CEP 70919-970 - Brasília - DF - Brasil

## **Agradecimentos**

*Agradeço a Deus, a minha fortaleza, por estar presente durante toda a minha vida e permitir a conclusão de mais esta etapa em minha carreira. À minha esposa Aurineia e meu filho Joab, pela motivação, compreensão e paciência nos últimos anos. Aos meus queridos pais Clara Cirino e Acrísio Gomes, por todo o apoio, mesmo estando distantes. Ao meu irmão Calby Cirino, por dar suporte em assuntos que tive de abrir mão para estudar. Ao meu orientador, professor Díbio Borges, pela paciência, pelo compartilhamento de conhecimento e pela brilhante condução da orientação. Ao colega Marcelo Brandão, pela parceria ao longo de todo o mestrado. À Universidade de Brasília, por meio do Decanato de Pós-graduação (DPG) e do Programa de Pós-graduação em Sistemas Mecatrônicos (PPMEC), por todo o suporte. Por fim, a todos os que de alguma forma contribuíram para a conclusão desta pesquisa.*

**JACÓ CIRINO GOMES**



---

## RESUMO

O reconhecimento rápido de pragas e insetos em lavouras é uma das estratégias de controle para mitigar perdas na produtividade agrícola. Algoritmos de aprendizagem de máquinas têm sido bem sucedidos como ferramentas de reconhecimento automático em aplicações envolvendo grandes conjuntos de dados, mas os modelos tendem a reduzir drasticamente seu desempenho quando lidam com conjuntos de dados pequenos, o que dificulta a ampla utilização dessa tecnologia em campo. A capacidade de aprender com poucas amostras visuais, é um dos principais desafios no campo da aprendizagem de máquinas. A *Few-shot Learning* (FSL), entretanto, é uma abordagem de aprendizagem recente e mais adequada para lidar com poucos dados. Neste trabalho, conduzimos um estudo sobre a FSL baseada em métrica para fornecer um modelo competitivo de classificação de insetos usando poucas amostras. O modelo proposto inclui dois módulos principais: 1) um mecanismo de redes neurais convolucionais (CNN) para extração e fusão de recursos multicamada (FMC) para a geração de vetores representativos ricos em informações de classes, e 2) um módulo de similaridade baseado nas divergências de Bregman para comparação de vetores gerados pela FMC. Dois conjuntos de dados desafiadores de insetos da agricultura foram propostos para os experimentos: o primeiro contendo insetos separados por estágio de maturidade, e o segundo abrangendo pragas e insetos benéficos (incluindo predadores e parasitoides) da cultura do milho. Uma ampla cadeia experimental foi realizada para análise de divergências e validação do modelo proposto. A validação do modelo foi realizada comparando-se os resultados aos de redes consolidadas na literatura. Os resultados mostraram que o modelo apresentado melhorou a acurácia em até 3,62% e 2,82% em tarefas de classificação de *1-shot* e *5-shot* respectivamente usando a FMC. Além disso, o modelo superou os resultados das redes ResNet50, VGG16 e MobileNetV2, com redução em até 99% da quantidade de parâmetros de aprendizagem. Este trabalho mostrou que a aprendizagem *few-shot* é uma abordagem relevante para pesquisas relacionadas a classificação de insetos na agricultura, servindo como uma alternativa para o desenvolvimento de soluções de prototipagem rápida para aplicação em campo.

**Palavras-chaves:** *Few-shot Learning*, classificação de insetos na agricultura, visão computacional, fusão de características.

---

# ABSTRACT

Rapid recognition of crop insect pests is one of the control strategies to mitigate losses in agricultural productivity. Machine Learning algorithms have been successful as automatic recognition tools in large data sets applications. But models tend to drastically reduce their performance when dealing with small data sets, which makes it difficult for the wide use of this technology in the crop field. The ability to learn with few visual samples is one of the main challenges in the field of machine learning. Few-shot Learning (FSL), however, is a recent and suitable learning approach to deal with few samples. In this work, we conducted a study to provide a competitive metric-based FSL insect classification model. The proposed model includes two main modules: 1) an image feature extraction and fusion mechanism (FMC) based on convolutional neural networks (CNN) to generate information-rich representative vectors of classes, and 2) a similarity module based on Bregman divergences to compare vectors generated by the FMC. Two challenging agricultural insect data sets were proposed for the experiments: the first containing insect pests separated by maturity stages, and the second covering pests and beneficial insects (including predators and parasitoids) of the maize crops. A wide experimental chain was carried out for the divergence analysis and validation of the proposed model. Validation was performed by comparing our results to those of networks consolidated in the literature. The results showed that the presented model improved the accuracy by up to 3.62% and 2.82% in *1-shot* and *5-shot* classification tasks, respectively, using FMC. Furthermore, the model outperformed the results of the ResNet50, VGG16, and Mobilenetv2 networks, with the advantage of reducing the number of learnable parameters by up to 99%. This work showed that Few-shot learning is a relevant approach for research related to insect classification in agriculture, serving as an alternative for the development of rapid prototyping solutions for field application.

**Keywords:** Few-shot learning, insect pest classification, computer vision, multi-scale feature fusion.

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
1.1	Contextualização . . . . .	1
1.2	Objetivos da Dissertação . . . . .	2
1.2.1	Objetivo geral . . . . .	2
1.2.2	Objetivos específicos . . . . .	3
1.3	Contribuições da Dissertação . . . . .	3
1.4	Organização do Documento . . . . .	4
<b>2</b>	<b>REVISÃO DA LITERATURA E FUNDAMENTAÇÃO TEÓRICA</b>	<b>5</b>
2.1	Reconhecimento de insetos na agricultura . . . . .	5
2.2	Aprendizagem <i>Few-shot</i> . . . . .	8
2.2.1	Contexto histórico . . . . .	8
2.2.2	Modelos baseados em métrica . . . . .	9
2.2.3	Aprimoramento dos modelos de última geração . . . . .	14
<b>3</b>	<b>ANÁLISE DE DIVERGÊNCIAS EM APRENDIZAGEM <i>Few-shot</i></b>	<b>17</b>
3.1	Divergências de Bregman . . . . .	17
3.1.1	Distância Euclidiana quadrática . . . . .	18
3.1.2	Distância quadrática de Mahalanobis . . . . .	18
3.1.3	Divergência de Kullback-Leibler . . . . .	18
3.1.4	Divergência de Itakura-saito . . . . .	19
3.2	O modelo adotado para análise de divergências . . . . .	19
3.3	O conjunto de dados IP-FSL . . . . .	22
3.4	Experimental . . . . .	24
3.5	Resultados e discussões . . . . .	25
3.5.1	Os efeitos da variação de <i>K-shot</i> . . . . .	26
3.5.2	Os efeitos da variação de <i>N-way</i> . . . . .	26
3.5.3	Análise das divergências de Bregman . . . . .	30
3.6	Conclusões do capítulo . . . . .	37
<b>4</b>	<b>APRENDIZAGEM <i>Few-shot</i> VIA FUSÃO DE RECURSOS E ENTROPIA RELATIVA</b>	<b>39</b>
4.1	O modelo proposto . . . . .	39
4.1.1	Extração de características de imagem . . . . .	41
4.1.2	Medição de similaridade . . . . .	44

4.2	Conjunto de dados de insetos do milho . . . . .	45
4.3	Detalhes experimentais . . . . .	46
4.4	Resultados e discussões . . . . .	48
4.5	Conclusões do capítulo . . . . .	52
<b>5</b>	<b>CONCLUSÕES</b>	<b>53</b>
5.1	Os conjuntos de dados propostos . . . . .	53
5.2	As divergências de Bregman em FSL . . . . .	53
5.3	Modelo de classificação proposto . . . . .	54
5.4	Sugestões de Trabalhos Futuros . . . . .	54
5.5	Artigos publicados . . . . .	55
	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>56</b>
	<b>APÊNDICES</b>	<b>61</b>
<b>A</b>	<b>INFORMAÇÕES DO CONJUNTO DE DADOS IP-FSL</b>	<b>62</b>
<b>B</b>	<b>RESUMO DO MODELO PROPOSTO PARA EXTRAÇÃO DE RECURSOS MULTICAMADA</b>	<b>63</b>

# LISTA DE FIGURAS

2.1	Um exemplo de conjuntos de meta-treinamento e meta-teste para classificação de domínio misto. Fonte: Adaptado de (LI; YANG, 2021) . . . . .	7
2.2	Um exemplo de meta-dados para classificação de domínio cruzado. Fonte: Adaptado de (LI; YANG, 2021) . . . . .	8
2.3	Conjuntos de dados em uma tarefa de classificação <i>few-shot</i> . . . . .	11
2.4	Estratégia da meta-aprendizagem em aprendizagem <i>few-shot</i> . A figura ilustra tarefas de classificação <i>2-way</i> e <i>2-shot</i> usando uma imagem de consulta $q$ . Em algumas abordagens, as redes de extração de recursos de imagens $f$ e $g$ são a mesma. As imagens de consulta são rotuladas de acordo com uma pontuação de similaridade com os <i>embeddings</i> de suporte. . . . .	12
2.5	Estrutura das redes Siamesas. Fonte: Adaptado de (KOCH et al., 2015) . . . . .	13
2.6	Esquemático das redes de Trigêmeos. Fonte: Adaptado de (HOFFER; AILON, 2015) . . . . .	14
2.7	Esquemático das redes <i>Matching</i> usando uma configuração <i>4-way</i> e <i>1-shot</i> , com 1 imagem de consulta $q$ . Fonte: (VINYALS et al., 2016) . . . . .	15
2.8	Estratégia das redes Prototípicas. Fonte: (SNELL; SWERSKY; ZEMEL, 2017) . . . . .	15
2.9	Rede de Relação demonstrada a partir de uma configuração <i>5-way</i> e <i>1-shot</i> , com 1 imagem de consulta $q$ . Fonte: Adaptado de (SUNG et al., 2018) . . . . .	16
3.1	Episódio de treinamento. . . . .	20
3.2	Configuração dos dados IP-FSL. . . . .	23
3.3	Algumas amostras dos dados IP-FSL. Os rótulos de classes estão numerados conforme os nomes dos insetos na Tabela A.1. A figura mostra alguns exemplos de classes que possuem insetos das duas fases, por isso aparecem nos dois subconjuntos (são eles, 2, 31, 67, 74, 87 e 89). . . . .	24
3.4	Tendência de acurácia em relação aos parâmetros <i>N-way-K-shot</i> para classificação de insetos da fase adulta. . . . .	27
3.5	Tendência de acurácia em relação aos parâmetros <i>N-way-K-shot</i> para classificação de insetos da fase primária. . . . .	28
3.6	Um exemplo da perda e acurácia de treinamento usando o conjunto de dados da fase adulta e a divergência KL em uma configuração de <i>5-shot</i> . . . . .	29
3.7	Comparação das divergências de Bregman para a fase adulta. ED: Distância Euclidiana. MD: Distância de Mahalanobis. KL: Divergência de Kullback-Leibler. IS: Divergência de Itakura-Saito. . . . .	31

3.8	Comparação das divergências de Bregman para a fase primária. ED: Distância Euclidiana. MD: Distância de Mahalanobis. KL: Divergência de Kullback-Leibler. IS: Divergência de Itakura-Saito. . . . .	32
3.9	Acurácia para classes similares de insetos adultos usando <i>10-shot</i> . . . . .	34
3.10	Média das acurácias usando as configurações mais promissoras para classificação <i>few-shot</i> . . . . .	35
3.11	Tempo gasto após um episódio de treinamento usando <i>3-way</i> , <i>10-shot</i> e $q = 5$ . . .	36
3.12	Taxa Acurácia-tempo. . . . .	37
4.1	Arquitetura do modelo proposto. . . . .	40
4.2	Visualização e entendimento das camadas da CNN. Fonte: Adaptado de (ZEILER; FERGUS, 2014). . . . .	42
4.3	O modelo proposto para extração e fusão de recursos multicamadas da CNN (FMC). . . . .	43
4.4	Exemplo representativo de extração e fusão de recursos multicamada (FMC) de uma imagem. . . . .	43
4.5	Desempenho esperado no cálculo da divergência. Resultados obtidos usando o modelo treinado. . . . .	45
4.6	Amostras de imagens de cada classe dos dados. . . . .	47
4.7	Tendência da acurácia a medida que mais recursos $f$ são fundidos. . . . .	50
4.8	Performance of the models. . . . .	50
4.9	Comparação dos melhores desempenhos dos <i>backbones</i> . . . . .	51
B.1	Resumo do modelo de extração de recursos multicamada, composto pelo <i>backbone</i> principal e suas ramificações $f_x$ ( $x = 1, 2, 3, 4$ ). . . . .	63

# LISTA DE TABELAS

2.1	Principais artigos de FSL voltados para classificação de insetos. . . . .	6
3.1	Resultados da classificação de insetos na fase adulta. . . . .	26
3.2	Resultados da classificação de insetos na fase primária. . . . .	26
4.1	Classes do conjunto de dados de insetos das culturas de milho. . . . .	46
4.2	Número de parâmetros em cada modelo. . . . .	47
4.3	Acurácia média de cada modelo. . . . .	49
A.1	Informações dos dados IP-FSL. Essa é uma versão compacta do conjunto de dados IP102 (WU et al., 2019) construída especificamente para aprendizagem <i>few-shot</i> . Foram mantidos os nomes dos insetos conforme publicado no artigo de referência, podendo conter nomes comuns, bem como científicos. . . . .	62

# LISTA DE ACROGRAMAS E ABREVIATURAS

## Subscritos

<i>test</i>	Teste
<i>train</i>	Treino
<i>n</i>	A classe <i>n</i>

## Siglas

ARM	<i>Advanced RISC Machine</i>
CNN	<i>Convolutional Neural Networks</i> - Redes Neurais Convolucionais
ED	<i>Euclidean distance</i> - Distância Euclidiana
FEAT	<i>Few-shot Embedding Adaptation with Transformer</i>
FMC	Fusão de recursos Multicamada
FN	<i>False Negative</i> - Falso Negativo
FP	<i>False Positive</i> - Falso Positivo
FPGA	<i>Field-Programmable Gate Array</i>
FSL	<i>Few-shot Learning</i> - Aprendizagem com poucos dados
GAN	<i>Generative Adversarial Networks</i> - Redes Contraditórias Generativas
IS	Itakura-Saito
IP-FSL	<i>Insect Pest - Few-shot Learning</i>
KL	Kullback-Leibler
L2M	<i>Learning to Measure</i>
MD	<i>Mahalanobis distance</i> - Distância de Mahalanobis
RGB	<i>Red, Green, and Blue</i> - Vermelho, Verde e Azul
SFR	Saída da estrutura de Fusão de Recursos
SGD	<i>Stochastic Gradient Descent</i> - Gradiente Descendente Estocástico
TP	<i>True Positive</i> - Verdadeiro Positivo
TN	<i>True Negative</i> - Verdadeiro Negativo

## Abreviaturas

<i>K-shot</i>	Quantidade de imagens de suporte em uma classe
<i>N-way</i>	Quantidade de classes em uma tarefa de classificação
ProtoNet	<i>Prototypical Networks</i> - Redes Prototípicas



# 1 INTRODUÇÃO

Este capítulo apresenta uma introdução sobre o contexto geral dessa pesquisa, descrevendo os objetivos, contribuições e a organização da dissertação.

## 1.1 CONTEXTUALIZAÇÃO

As pragas agrícolas são organismos caracterizados pelo aumento da população de insetos causadores de danos às lavouras, sendo responsáveis por perdas importantes na agricultura global (OLIVEIRA et al., 2014; THENMOZHI; REDDY, 2019). Por esse motivo, o gerenciamento de infestações demanda reconhecimento rápido de insetos para tomada de decisão e redução de custos na produção agrícola.

O reconhecimento de insetos em plantações é uma das tarefas mais desafiadoras na agricultura devido à existência de muitas espécies na natureza e da semelhança entre muitas delas. Por essa razão, diferenciar insetos com acurácia para o controle de infestações não é uma solução trivial. A classificação automática de insetos, nesse contexto, é vista como uma solução potencialmente promissora no combate às infestações.

A aprendizagem profunda (*Deep learning*) tem sido vastamente utilizada como solução de última geração para reconhecimento visual de insetos na agricultura (LI et al., 2020; KASINATHAN; SINGARAJU; UYYALA, 2021). No entanto, os algoritmos demandam conjuntos de dados tipicamente grandes para treinamento, além de custo computacional elevado. Tendo em vista que a construção de conjuntos de dados de insetos é naturalmente onerosa do ponto de vista da captura em condições reais de campo, a demanda por algoritmos de aprendizagem de máquinas que lidam com poucos dados é crescente.

A capacidade humana de aprender com dados limitados ainda supera os algoritmos clássicos existentes em Aprendizagem de Máquinas. Por outro lado, a *Few-shot learning* (FSL) (WANG et al., 2020) é uma técnica que define uma aproximação relativa entre o aprendizado de máquinas e o aprendizado humano, considerando a tarefa desafiadora de aprender a partir de poucas amostras. A FSL pode fornecer uma alternativa de prototipagem rápida e útil para aplicações na agricultura, servindo como facilitadora no reconhecimento de espécies de insetos usando poucas amostras de dados ou com classes naturalmente desbalanceadas na natureza.

O conceito de FSL teve os primeiros registros nos anos 2000 (LU et al., 2020). Apesar disso, essa abordagem ganhou maior projeção com o advento da aprendizagem profunda e visão computacional, sobretudo, com o crescente movimento das Redes Neurais Convolucionais (CNN) a partir de 2015. Atualmente, há registros de aplicação da FSL em diversos seguimentos da aprendizagem de máquinas com resultados relevantes, tais como, na classificação de imagens

(DHILLON et al., 2020), detecção de objetos (KANG et al., 2019), bem como no processamento de linguagem natural (PENG et al., 2020).

Um algoritmo clássico de FSL consiste na extração de características de imagens a partir de uma CNN (*encoder*) para classificação baseada na similaridade entre imagens. Ao longo dos últimos anos, diversos métodos foram propostos para melhoria no desempenho dos modelos. Embora muitos avanços tenham sido alcançados, duas características são muito comuns na maioria dos trabalhos, como o uso de redes muito complexas resultantes da combinação de vários métodos para extração de recursos de imagens mais representativos, e o uso de métodos recorrentes para medição de similaridade entre imagens.

Dado que na aprendizagem *few-shot* poucas amostras são disponibilizadas ao modelo para a construção do aprendizado, a extração de recursos de imagens para formação de *embeddings* ricos em informações é uma tarefa crucial para o desempenho final do modelo. Entretanto, manter uma arquitetura simples para extração de recursos relevantes é desafiador. Por isso, o aprimoramento e a otimização das técnicas de FSL para classificação automática são necessários para melhoria dos resultados alcançados, aumentando as possibilidades com respeito ao uso dessa abordagem em aplicações reais de campo.

A classificação de insetos na agricultura requer métodos robustos para a diferenciação de imagens em diferentes situações, sobretudo, para lidar com classes de insetos muito semelhantes, a partir de poucas amostras disponíveis. Nesta dissertação, propomos abordar a classificação de insetos na agricultura a partir de um modelo aprimorado de FSL baseado na eficiência e com baixa complexidade em relação aos métodos de última geração utilizados para tal.

Esta dissertação propõe um modelo de classificação *few-shot* para reconhecimento de insetos importantes na agricultura, construído para lidar com os desafios da classificação mesmo com poucos dados disponíveis para treinamento. O modelo proposto será testado em aplicações específicas de classificação de insetos a partir de dados exclusivos, considerando o reconhecimento de insetos por fases de maturidade (GOMES; BORGES, 2022), com potencial para evitar precocemente as infestações, bem como pela diferenciação de pragas e insetos benéficos (incluindo predadores e parasitoides), com potencial para o controle biológico de pragas.

## **1.2 OBJETIVOS DA DISSERTAÇÃO**

### **1.2.1 Objetivo geral**

Propor um modelo computacional aprimorado de *Few-shot learning* para classificação de insetos importantes na agricultura usando poucas amostras de dados.

## 1.2.2 Objetivos específicos

São objetivos específicos dessa dissertação:

- Realizar a construção de conjuntos de dados de insetos importantes na agricultura para aprendizagem *few-shot*.
- Realizar um estudo comparativo das divergências de Bregman como medidas de similaridade em um modelo de FSL baseado em métrica.
- Investigar a influência dos parâmetros *N-way-K-shot* do modelo para a definição de caminhos promissores relacionados às perspectivas de acertos na classificação *few-shot*.
- Propor um mecanismo CNN de extração e fusão de recursos de baixo nível (locais) e de alto nível (globais) de imagens baseado em multicamadas (FMC) para a geração de vetores (*embeddings*) com forte representatividade de classe.
- Realizar testes de classificação *few-shot* de insetos em duas tarefas desafiadoras em agricultura: a classificação baseada no ciclo de vida do inseto e, para diferenciar insetos entre benéficos e pragas da cultura do milho.

## 1.3 CONTRIBUIÇÕES DA DISSERTAÇÃO

Esta Dissertação fornece três importantes contribuições para classificação *few-shot* de insetos na agricultura, as quais serão apresentadas contextualmente a seguir:

- Apesar da notória relevância das divergências de Bregman na comparação de distribuições, seus métodos não foram amplamente investigados em FSL. Este trabalho fornece um estudo exclusivo sobre a comparação de quatro importantes divergências de Bregman para classificação de insetos usando aprendizagem *few-shot*, são elas: as distâncias Euclidiana e de Mahalanobis, e as divergências de Kullback-Leibler e Itakura-Saito. A adoção de uma divergência adequada é importante para a diferenciação de imagens de insetos com acurácia devido à semelhança entre muitas espécies. O estudo serve como um embasamento para a escolha da divergência em futuras pesquisas.
- Este trabalho propõe um modelo de FSL baseado na fusão de características de imagens e Entropia relativa para classificação de insetos com acurácia relevante. A fusão multicamadas (FMC) proporciona a geração de vetores com maior representatividade de classe, comparando-se aos métodos que usam apenas recursos globais de imagens existentes na literatura. Já a entropia relativa consegue comparar esses vetores com maior eficiência em vez da distância Euclidiana, comumente usada em FSL, em operações de classificação 1-shot. O modelo proposto pode ser adaptado para outras aplicações envolvendo classificação visual.

- Este trabalho propõe dois conjuntos de dados de insetos importantes na agricultura para aprendizagem *few-shot* supervisionada. O primeiro conjunto contém insetos com categorias separadas por estágios de maturidade (fases primária e adulta), usado aqui para análise de divergências. O segundo com separação de insetos benéficos e pragas da cultura do milho, usado aqui para teste do modelo proposto nessa pesquisa.

## 1.4 ORGANIZAÇÃO DO DOCUMENTO

O restante deste documento está organizado como segue:

- O Capítulo 2 apresenta uma revisão da literatura e a fundamentação teórica que envolve a aprendizagem *Few-shot* baseada em métrica.
- O Capítulo 3 fornece um estudo sobre as divergências de Bregman aplicadas em FSL, esclarecendo as situações, bem como as configurações dos parâmetros *N-way-K-shot*, que fornecem as melhores possibilidades de acertos na classificação.
- O Capítulo 4 propõe um modelo de aprendizagem *few-shot* aprimorado, com aplicação na diferenciação entre insetos benéficos e pragas do milho.
- As conclusões e sugestões de trabalhos futuros são apresentadas no Capítulo 5.

## 2 REVISÃO DA LITERATURA E FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta uma visão geral sobre a classificação automática de insetos na agricultura. Além disso, uma seção apresenta os conceitos que envolvem a aprendizagem *few-shot*, cobrindo o contexto histórico e a fundamentação teórica dos modelos baseados em métrica, além dos principais aprimoramentos da FSL publicados nos últimos anos.

### 2.1 RECONHECIMENTO DE INSETOS NA AGRICULTURA

Entre os problemas agrícolas mais investigados envolvendo visão computacional e aprendizagem de máquinas, a classificação de insetos e doenças de plantas são os fatores que receberam maior atenção (YANG et al., 2022). A identificação rápida de agentes patológicos em plantações é crucial para o controle da produtividade agrícola. Entretanto, o reconhecimento de insetos é um caso naturalmente complexo, visto que existe um número expressivo de espécies na natureza. Além disso, a semelhança entre insetos torna o reconhecimento automático uma das tarefas mais desafiadoras da agricultura.

O desenvolvimento de algoritmos de aprendizagem de máquinas baseados em imagens, tem contribuído fortemente para os processos de reconhecimento automático de insetos. As técnicas de aprendizagem profunda são majoritariamente usadas para o reconhecimento de pragas nas lavouras (ALVES; SOUZA; BORGES, 2020; LIU; WANG, 2020). No entanto, esses modelos normalmente requerem alto custo computacional e uma grande massa de dados para treinamento, o que nem sempre está disponível. As técnicas como o aumento de dados (KUSRINI et al., 2020) e transferência de aprendizado (THENMOZHI; REDDY, 2019) foram abordagens comumente adotadas para lidar com a falta de dados em muitos casos.

Embora as técnicas citadas acima tenham contribuído para o alcance de resultados promissores em aprendizagem profunda, há uma crescente necessidade de reconhecimento rápido de insetos a partir de algoritmos otimizados, com menor dependência de *hardwares* potentes, e de grandes conjuntos de dados genéricos como o ImageNet (RUSSAKOVSKY et al., 2015). Trabalhos anteriores mostraram um crescimento na tendência por algoritmos otimizados para classificação de insetos com aplicação *mobile* (NGUGI; ABELWAHAB; ABO-ZAHHAD, 2021), bem como por modelos com capacidade relevante de extração de características de imagens usando CNN rasa (LI; NIE; CHAO, 2020).

A aprendizagem *few-shot*, portanto, tem sido uma alternativa promissora para abordar questões desafiadoras em agricultura, como a classificação de patologias vegetais e de pragas. Trabalhos

recentes mostraram a grande importância dessa abordagem no reconhecimento de doenças foliares (LI; CHAO, 2021; ARGÜESO et al., 2020). Por outro lado, muitas doenças de plantas são causadas por agentes transmissores como as pragas (THENMOZHI; REDDY, 2019). Além disso, esses agentes são responsáveis por perdas significativas na agricultura global (OLIVEIRA et al., 2014; LI et al., 2020). Por essas razões, a identificação precoce desses agentes é uma alternativa promissora para mitigar os efeitos negativos das infestações em culturas. De fato, estudos focados em soluções para essa aplicação têm sido encontrados na literatura. A Tabela 2.1 mostra um panorama dos principais estudos voltados para a classificação de insetos usando FSL.

Tabela 2.1: Principais artigos de FSL voltados para classificação de insetos.

Autor	Título	Aplicação de classificação
(LI; YANG, 2020)	<i>Few-shot cotton pest recognition and terminal realization</i>	Insetos
(LI; CHAO, 2020)	<i>ANN-based continual classification in agriculture</i>	Insetos e doenças de plantas
(YANG et al., 2021)	<i>Small-sample learning with salient-region detection and center neighbor loss for insect recognition in real-world complex scenarios</i>	Insetos
(LI; YANG, 2021)	<i>Meta-learning baselines and database for few-shot classification in agriculture</i>	Insetos e doenças de plantas
(NUTHALAPATI; TUNGA, 2021)	<i>Multi-Domain Few-Shot Learning and Dataset for Agricultural Applications</i>	Insetos e doenças de plantas

A estrutura otimizada dos modelos de FSL facilita o desenvolvimento de soluções embarcadas, e o uso dessa tecnologia em aplicações remotas. Li e Yang (2020), por exemplo, usaram FSL para a classificação de pragas da cultura de algodão em um terminal embarcado. As redes prototípicas foram implementadas na placa de desenvolvimento PYNQ-Z2, a qual possui um *chip* integrado composto por um controlador ARM e uma estrutura FPGA responsável pelo cálculo da rede convolucional. Já Yang et al. (2021) apostaram em melhorias nas redes Prototípicas, propondo um mecanismo de detecção de regiões salientes para classificação de insetos. O modelo foi treinado para localizar regiões com características discriminatórias mais relevantes de insetos em imagens.

A meta-aprendizagem, usada na maioria dos trabalhos recentes, tem sido útil para classificação *few-shot* de múltiplos domínios em agricultura. Li e Yang (2021), por exemplo, investigaram a classificação de insetos e de doenças foliares em um problema de classificação de domínio cruzado (*cross-domain*). Os resultados mostraram que o domínio misto, no qual o meta-treinamento e o meta-teste usam classes de ambas as categorias de dados juntas (insetos e plantas), produziu melhores resultados. Nuthalapati e Tunga (2021) melhoraram esses resultados usando um mecanismo de transformação dos *embeddings* com uma estrutura de Transformer. As figuras 2.1 e 2.2 mostram exemplos de domínios misto e cruzado respectivamente.

Embora a meta-aprendizagem forneça uma metodologia promissora para a classificação de domínio cruzado, esta é uma tarefa desafiadora devido à limitação dos modelos com respeito ao acúmulo de conhecimento. Como solução para este problema, Li e Chao (2020) propuseram a classificação contínua de doenças foliares e insetos por meio do armazenamento e recuperação de informações de tarefas anteriores. As redes Generativas (GAN) foram usadas para gerar

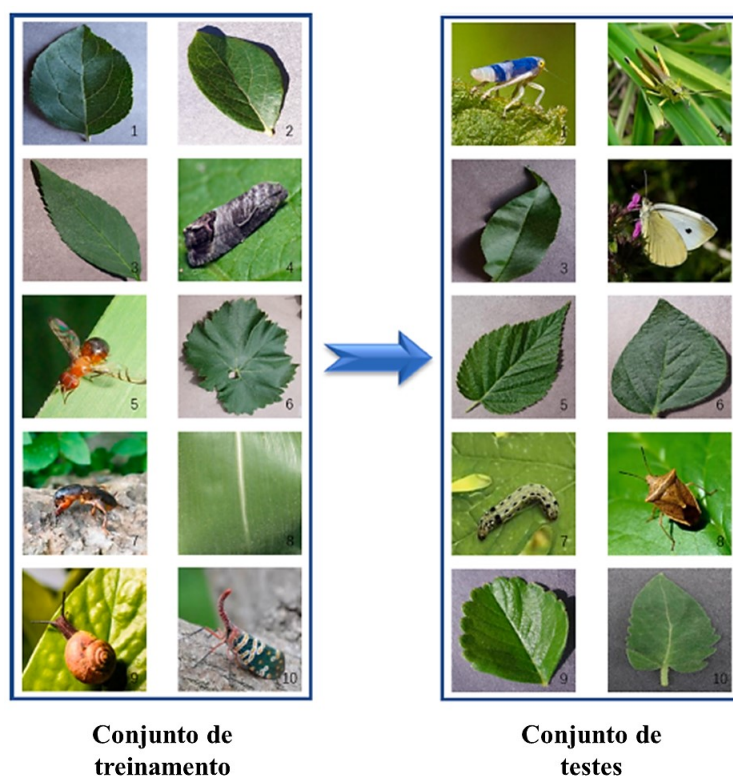


Figura 2.1: Um exemplo de conjuntos de meta-treinamento e meta-teste para classificação de domínio misto.  
 Fonte: Adaptado de (LI; YANG, 2021)

informações das tarefas. Essas informações são armazenadas e posteriormente utilizadas em novas tarefas de classificação.

Estudos recentes mostraram que a FSL é uma alternativa relevante para reconhecimento automático de insetos, apresentando resultados comparáveis aos de modelos profundos (SNELL; SWERSKY; ZEMEL, 2017). Ao menos três vantagens podem ser listadas sobre a aplicação da FSL na classificação de insetos da agricultura:

- *A necessidade de poucos dados* – algumas amostras de imagens são suficientes para treinamento do modelo, reduzindo significativamente o tempo gasto na coleta e anotação de imagens para construção de bancos de dados.
- *O uso de modelos otimizados* – existe a possibilidade de utilização de uma rede CNN rasa para extração de recursos de imagens, resultando em baixo custo computacional durante o treinamento e teste.
- *Aplicações embarcadas* – a viabilidade econômica em virtude da possibilidade de implementação em dispositivos com poder computacional limitado, tornando o método mais abrangente e acessível.

Muitos esforços foram empenhados nos últimos anos na busca de elevar os resultados em FSL. Na próxima seção, são apresentadas algumas considerações importantes sobre o contexto

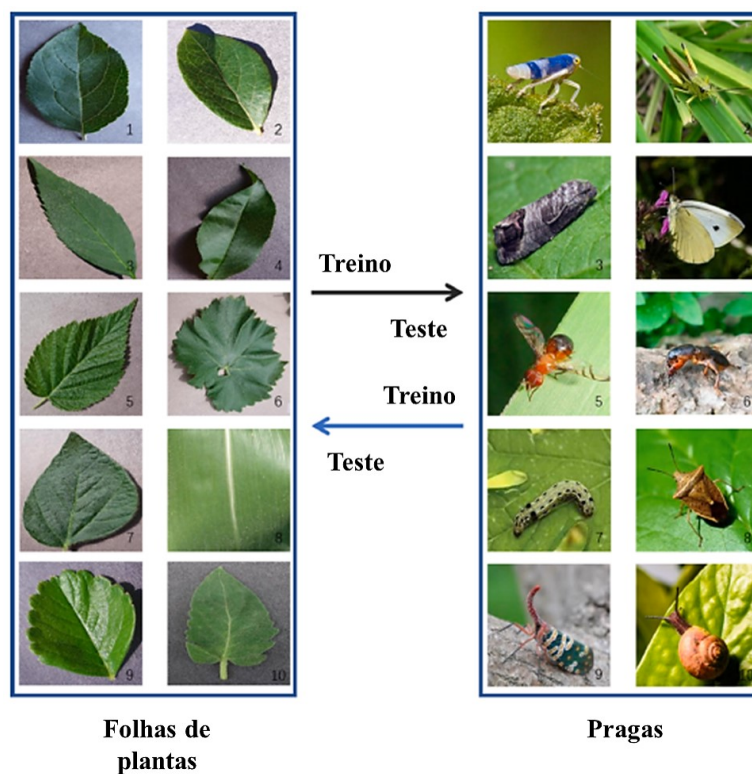


Figura 2.2: Um exemplo de meta-dados para classificação de domínio cruzado.  
 Fonte: Adaptado de (LI; YANG, 2021)

histórico e as principais técnicas de FSL, bem como as melhorias mais importantes nos modelos de última geração, dando ênfase nos esclarecimentos sobre como o modelo aqui apresentado contribui com esses trabalhos.

## 2.2 APRENDIZAGEM *FEW-SHOT*

### 2.2.1 Contexto histórico

A aprendizagem *Few-shot* é um conceito relativamente recente na literatura, proposta inicialmente no início dos anos 2000 (LU et al., 2020). O regime geral dessa abordagem foi marcado por dois períodos importantes: o período da aprendizagem não profunda e o período profundo. No primeiro período, os primeiros trabalhos focaram na compreensão do aprendizado humano para o desenvolvimento de algoritmos com habilidade de generalização a partir de poucas amostras. Miller, Matsakis e Viola (2000) publicaram um dos estudos pioneiros nesse âmbito. Pouco tempo depois, o conceito de *one-shot learning* começou a ser articulado por Fe-Fei et al. (2003). No período da aprendizagem profunda, a FSL foi inicialmente impulsionada por Koch et al. (2015), que propuseram o primeiro estudo combinando CNN e FSL.

Apesar da relevância do aprendizado usando poucos dados, poucos estudos focaram no con-



ceito de FSL durante o período inicial para solução de problemas em geral. No entanto, essa abordagem ganhou maior projeção a partir dos modelos discriminativos, viabilizados pelo desenvolvimento da aprendizagem profunda. O poder de extração de características de imagens proporcionado pelas CNNs causou um forte impacto nos modelos de *few-shot*. A combinação do aprendizado profundo com o FSL acelerou o desenvolvimento de novas abordagens em poucos anos, culminando no aumento do número de estudos publicados a partir de 2015.

Os primeiros trabalhos envolvendo a combinação de CNN e FSL propuseram classificadores baseados na comparação de imagens por pares. Como o próprio nome sugere, tais comparadores tinham a função de julgar se um par de imagens pertencia a uma mesma classe ou se dividia em classes distintas. Koch et al. (2015) propuseram um modelo de redes Siamesas para a classificação de imagens a partir de tarefas *one-shot*. A partir desse estudo, novos modelos baseados em classificação multiclasse foram surgindo. Em 2016, Vinyals et al. (2016) propuseram as redes *Matching*. Nos anos seguintes, foram publicadas as *Prototypical networks* (SNELL; SWERSKY; ZEMEL, 2017) e *Relation networks* (SUNG et al., 2018).

As abordagens iniciais do período profundo, bem como as suas variações subsequentes, por exemplo, (ORESHKIN; LÓPEZ; LACOSTE, 2018; YE et al., 2020; LIU et al., 2020; WANG; WANG, 2021), aproveitaram amplamente as vantagens da CNN para extração de recursos representativos de imagens, potencializando as pesquisas e levando o conceito de FSL para um novo período de desenvolvimento rápido. As técnicas adotadas se concentraram na otimização dos modelos profundos para lidar com problemas de FSL sob diferentes aspectos, incluindo a aprendizagem métrica (KAYA; BILGE, 2019) e meta-aprendizagem (HOCHREITER; YOUNGER; CONWELL, 2001), permanecendo até agora como abordagens de última geração em FSL.

A evolução da FSL mostrou que a transição do período não profundo para o profundo causou transformações relevantes na arquitetura dos modelos. A combinação de CNN, para extração de recursos de imagens, com a aprendizagem métrica e a meta-aprendizagem, é uma tática comumente adotada na maioria dos trabalhos. Atualmente, a FSL aparece com frequência em pesquisas que envolvem visão computacional, atraindo uma ampla atenção da comunidade de aprendizagem de máquinas. A próxima seção apresenta mais detalhes sobre as abordagens que usaram a aprendizagem métrica em conjunto com a meta-aprendizagem.

### **2.2.2 Modelos baseados em métrica**

Os métodos de FSL baseados em métrica (*learninf-to-measure* - L2M) (LU et al., 2020), consistem na obtenção de conhecimento a partir da semelhança entre imagens. Um modelo típico de L2M é composto por um extrator CNN de recursos de imagem e uma função de semelhança para diferenciar imagens no espaço de características. Alguns dos principais diferenciais dos modelos baseados em métrica dizem respeito à simplicidade e eficiência na comparação de características para generalização de imagens, mesmo a partir de poucos exemplos de treinamento.

Se os dados de um determinado problema forem insuficientes para o treinamento dos modelos

de aprendizagem, a obtenção de experiência a partir de problemas com dados semelhantes pode ser vista como uma solução. Por esse motivo, a maioria das abordagens caracteriza a aprendizagem *few-shot* como um problema de meta-aprendizagem. A combinação da meta-aprendizagem com o L2M, tem sido usada com frequência em modelos FSL de última geração como estratégia de aprendizagem em domínio cruzado (*cross-domain*) (LI; YANG, 2021; JAMAL; QI, 2019; SUN et al., 2019).

A proposta da meta-aprendizagem se concentra no paradigma de aprender a aprender (*learn to learn*) diferindo da aprendizagem clássica. Na aprendizagem clássica, os modelos aprendem a generalizar a partir de dados de treinamento, em seguida o modelo é avaliado usando dados de teste. Normalmente, os modelos necessitam ser treinados novamente quando uma nova classe é inserida. Na meta-aprendizagem, o modelo aprende a aprender dado um conjunto de tarefas de treinamento usando dados de um problema específico. A avaliação do modelo ocorre em um conjunto de tarefas de testes para solução de outro problema, que pode ou não ter relação com os dados do problema usado no treinamento. Dessa forma, o modelo não necessita de um novo treinamento quando novas classes são inseridas para testes.

A rotina de treinamento dos modelos de FSL é composta por diversas tarefas de classificação baseadas no paradigma *N-way, K-shot*. O parâmetro  $N$  representa a quantidade de classes e  $K$  a quantidade de amostras de suporte em cada classe. Além disso, o parâmetro implícito,  $q$ , indica a quantidade de exemplos de consulta por classe. Diferentemente da aprendizagem clássica, em que o modelo normalmente necessita de um volume maior de dados, o conceito de aprendizagem *few-shot* considera uma quantidade pequena de exemplos de treinamento por classe. Por esse motivo, o parâmetro  $K$  varia tipicamente entre 1 e 10. Já o parâmetro  $N$  ( $N_{train}/N_{test}$ ) pode variar conforme o modelo ou a prática experimental, adotados.

O conjunto de imagens de suporte e de consulta dentro de uma determinada tarefa são representados por  $S = N \times K$  e  $Q = N \times q$  respectivamente. O objetivo do modelo é rotular as imagens do conjunto de consulta  $Q$  de acordo com  $N$  classes de uma tarefa, a partir de uma medida de similaridade. A Figura 2.3 ilustra os dados de uma tarefa de classificação específica.

O modelo é treinado usando um conjunto de origem (*source set*) dividido entre imagens de suporte (*trainx*) e de consulta (*trainy*). As imagens de consulta são utilizadas para avaliar o desempenho de treinamento do modelo em cada tarefa de classificação. Assim como no treinamento, os testes são realizados por tarefas de classificação, dessa vez usando um conjunto de destino, ou conjunto alvo (*target set*), também dividido entre imagens de suporte (*testx*) e de consulta (*testy*), de onde é obtida a acurácia média de classificação do modelo *few-shot*. A Figura 2.4 mostra a estratégia de treinamento e teste usada na meta-aprendizagem.

Em geral, a maioria dos algoritmos de FSL adotam a metodologia de meta-aprendizagem para treinamento dos modelos, garantindo um bom desempenho na classificação não restrita somente a novas imagens de classes já treinadas, mas também abrangendo a classificação de novas classes não treinadas, isto é, servindo como uma espécie de transferência de aprendizado. A metodologia de transferência de aprendizado usada na meta-aprendizagem garante ao modelo uma forma de

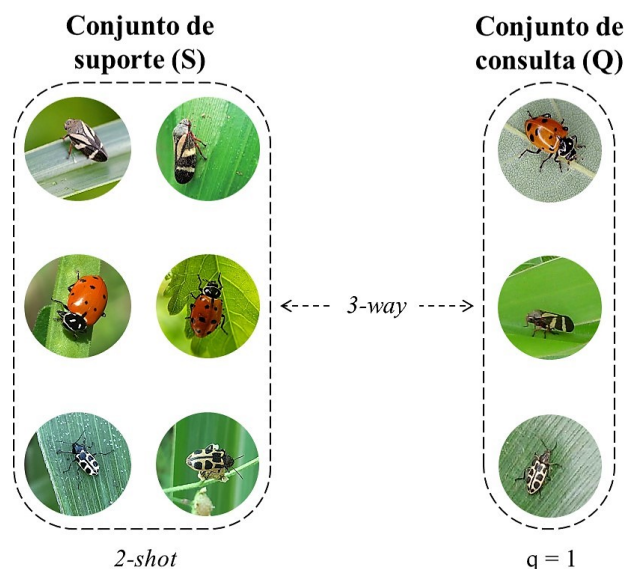


Figura 2.3: Conjuntos de dados em uma tarefa de classificação *few-shot*.

se adaptar a novas tarefas de classificação não vistas durante o treinamento. Nesse contexto, diferentemente da aprendizagem clássica, não há relação entre as classes contidas nos dados de treino (*source set*) e aquelas nos dados de teste (*target set*).

Os métodos de aprendizagem métrica baseados em meta-aprendizagem, formam um campo muito importante em FSL. Os algoritmos existentes são conceitualmente semelhantes e se diferem ao nível de implementação. Os modelos aprendem no espaço de características de imagens durante as tarefas de treinamento para diferenciar classes usando uma métrica de similaridade. A próxima seção apresenta uma breve descrição das principais arquiteturas baseadas em métrica.

### 2.2.2.1 Principais arquiteturas L2M usadas em meta-aprendizagem

#### Redes Siamesas (*Siamese Networks*)

Os primeiros trabalhos envolvendo FSL e redes baseadas em métrica focaram em comparadores por pares. Nas redes Siamesas, os pares de imagens passam pela mesma estrutura de rede para extração de recursos e o modelo aprende usando a Diferença Absoluta ( $L_1$ ). A saída do modelo é a probabilidade de duas imagens pertencerem a mesma classe. O esquema das redes Siamesas é apresentado na Figura 2.5.

Embora não seja explicitamente notável que os comparadores por pares possam ser usados como modelos de FSL, em virtude da estrutura padrão  $N$ -way,  $K$ -shot exigida, é possível considerar os pares de dados como sendo exemplos de suporte e consulta respectivamente. Dessa forma, a comparação entre cada par se torna uma tarefa de classificação. Koch et al. (2015) por exemplo, relataram resultados promissores em operação de *one-shot* usando as redes Siamesas. O modelo é treinado por várias tarefas de classificação usando diversos pares aleatórios de imagens. A partir do aprendizado adquirido, consegue generalizar em testes usando novas classes de dados.

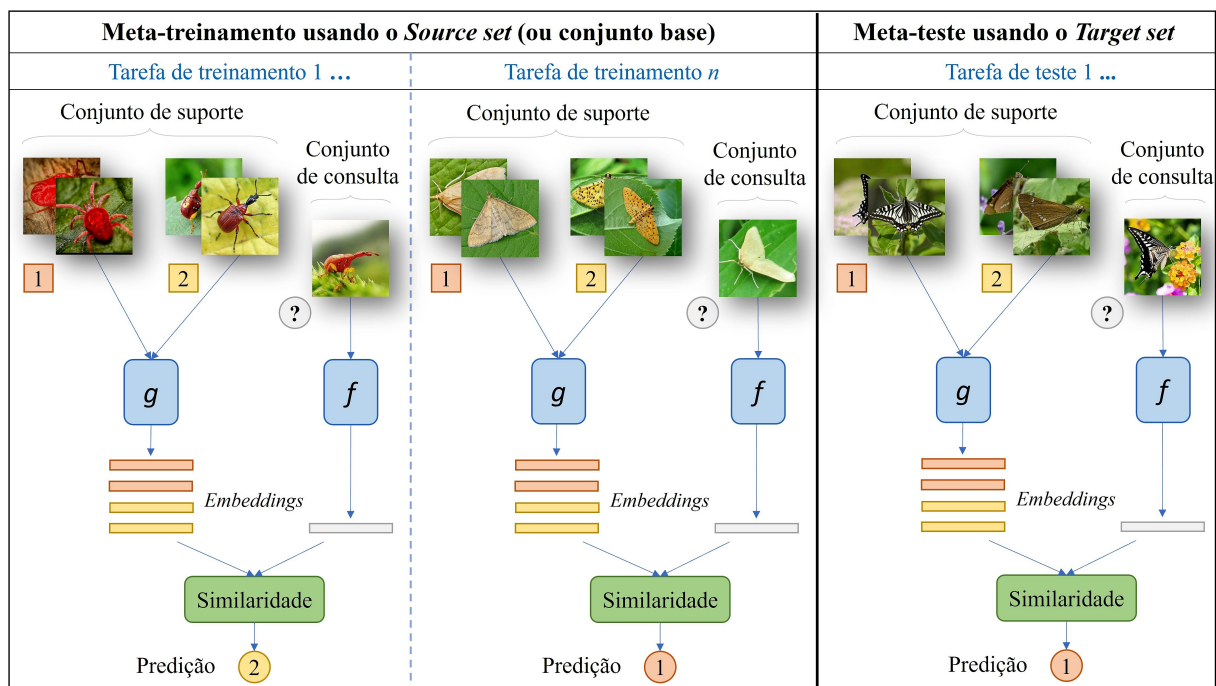


Figura 2.4: Estratégia da meta-aprendizagem em aprendizagem *few-shot*. A figura ilustra tarefas de classificação *2-way* e *2-shot* usando uma imagem de consulta  $q$ . Em algumas abordagens, as redes de extração de recursos de imagens  $f$  e  $g$  são a mesma. As imagens de consulta são rotuladas de acordo com uma pontuação de similaridade com os *embeddings* de suporte.

### Redes de Trigêmeos (*Triplet Networks*)

As redes de Trigêmeos (HOFFER; AILON, 2015) funcionam de maneira semelhante às redes Siamesas. A diferença é que ela considera três imagens em uma tarefa de classificação: a imagem  $x$ , um exemplo positivo  $x^+$  pertencente à mesma classe de  $x$ , e um exemplo negativo  $x^-$  de classe diferente. Os três exemplos passam pela mesma estrutura de rede CNN (*Net*) para extração de recursos de imagens, conforme mostra a Figura 2.6. Em vez de uma única comparação, ocorre um par de comparações entre  $(x, x^+)$  e  $(x, x^-)$ . A rede adota o critério de perda de trigêmeo para o treinamento do modelo, com o intuito de aproximar as características entre as imagens da mesma classe e distanciar em caso de classes diferentes.

### Redes *Matching* (*Matching Networks*)

As redes *Matching* (VINYALS et al., 2016) são exemplos de classificadores multiclasse e diferem das redes Siamesas sob dois aspectos principais: 1) as imagens de suporte e consulta passam por estruturas  $f_\theta$  e  $g_\theta$  respectivamente para extração de recursos (podendo ser potencialmente consideradas  $f_\theta = g_\theta$ ), conforme mostra a Figura 2.7, 2) a similaridade é obtida por meio da distância do cosseno entre os *embeddings* em vez da distância absoluta.

### Redes Prototípicas (*Prototypical Networks*)

A ideia principal das redes Prototípicas (SNELL; SWERSKY; ZEMEL, 2017) se concentra no fato de que a média dos *embeddings* de suporte produz representantes de classes ideais quando uma divergência Bregman, como a distância Euclidiana, é usada. O objetivo do modelo é ba-

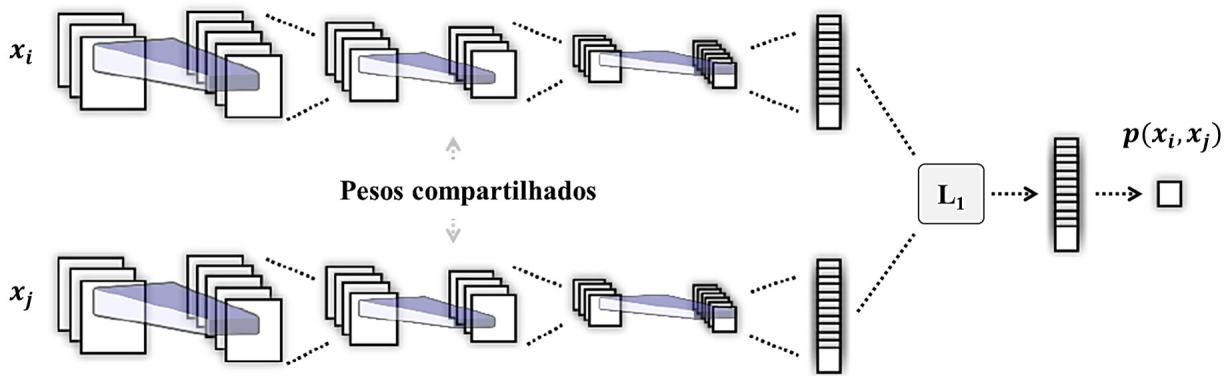


Figura 2.5: Estrutura das redes Siamesas.  
 Fonte: Adaptado de (KOCH et al., 2015)

sicamente aprender uma métrica no espaço de características de imagens que representa uma similaridade por distância, para a previsão de uma imagem com base no centróide (protótipo) de classe. As imagens de consulta são rotuladas encontrando o protótipo de classe mais próximo conforme mostra a Figura 2.8. Mais detalhes sobre o funcionamento das redes Prototípicas são apresentados no Capítulo 3.

### Redes de relação (*Relation Networks*)

As redes de Relação (SUNG et al., 2018) concatenam mapas de recursos das imagens de suporte aos de consulta e os envia para um módulo de relação para definição da similaridade. A Figura 2.9 mostra a arquitetura para a configuração *5-way* e *1-shot* com uma imagem de consulta  $q$ . O módulo de relação, constituído de camadas convolucionais e camadas totalmente conectadas, é responsável pela produção da respectiva pontuação de relação entre as imagens. Tal pontuação descreve a semelhança entre as imagens de suporte e consulta, em que zero indica baixo grau de similaridade e 1 significa similaridade alta.

Embora as redes citadas acima apresentem resultados diferentes entre diversas aplicações em relação à acurácia, elas possuem uma grande relação entre si, e não é possível especificar, em geral, qual o melhor modelo. A análise para adoção do modelo adequado depende estritamente da aplicação. As redes Siamesas e de Trigêmeos podem facilmente generalizar entre novas classes não treinadas em tarefas *2-way*. No entanto, em algumas aplicações que envolvem várias classes simultâneas, o uso das redes Prototípicas, *Matching* e *Relation* podem ser mais relevantes.

Em resumo, os algoritmos de meta-aprendizagem visam aprender a classificar por meio de tarefas para se adaptarem a novas tarefas. A meta-aprendizagem adéqua-se muito bem às primícias que envolvem a metodologia de FSL, sobretudo referindo-se aos modelos baseados em métrica. Estudos recentes mostraram os resultados promissores com o aprimoramento dessa abordagem, conforme relatado na seção seguinte.

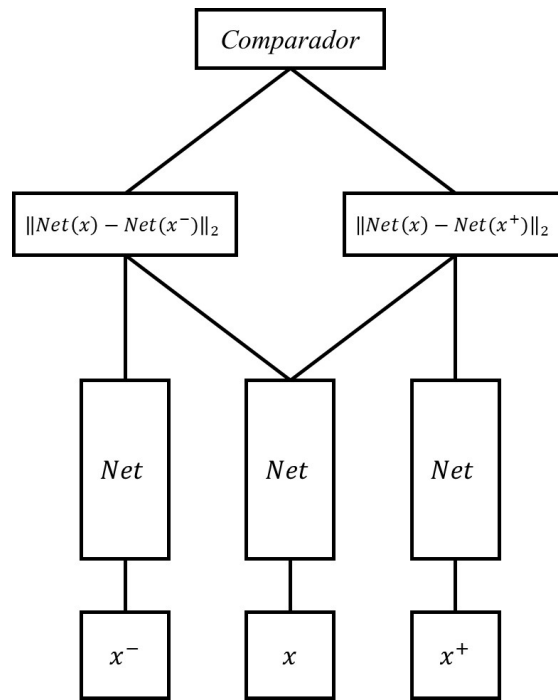


Figura 2.6: Esquemático das redes de Trigêmeos.  
 Fonte: Adaptado de (HOFFER; AILON, 2015)

### 2.2.3 Aprimoramento dos modelos de última geração

As buscas por algoritmos de aprendizagem baseados em poucos dados cresceram expressivamente a partir das mudanças na estrutura dos modelos de FSL, em 2015. Duas questões iniciais podem fazer muito sentido para o aprimoramento dessas abordagens: 1) como melhorar a extração de recursos discriminativos de imagens para formar *embeddings* com maior representatividade de classe e, 2) como obter maior desempenho na medição de similaridade entre *embeddings*.

Trabalhos recentes focaram na melhoria dos algoritmos para a extração de recursos de imagens (*feature extraction*). Ye et al. (2020) introduziram o FEAT para adaptar *embeddings* através de uma função específica (*set-to-set*) e uma arquitetura de Transformer. O FEAT melhora a extração de recursos discriminatórios enquanto produz *embeddings* específicos de tarefas. Ji et al. (2020) propuseram um modelo de redes Prototípicas baseado em um módulo de distribuição de peso sobre as instâncias para aumentar a representatividade dos protótipos. As imagens com mais informações intraclasse são associadas a um maior peso. Lin e Gao (2020) aprimoraram as redes Prototípicas propondo um extrator de recursos visuais que aprende características úteis de imagens reduzindo a quantidade de característica inúteis para geração dos *embeddings*. Já Yang et al. (2021) melhoraram as redes Prototípicas combinando reconhecimento e localização de objetos. Eles propuseram a detecção de região saliente para classificação de insetos em imagens com fundo complexo. Lim et al. (2021) aproveitaram os benefícios da transferência de aprendizado para a obtenção de *embeddings* de amostras mais consistentes usando as redes Prototípicas. O extrator de recursos de imagens usado foi pré-treinado usando outro conjunto de dados em uma tarefa não associada à aprendizagem *few-shot*.

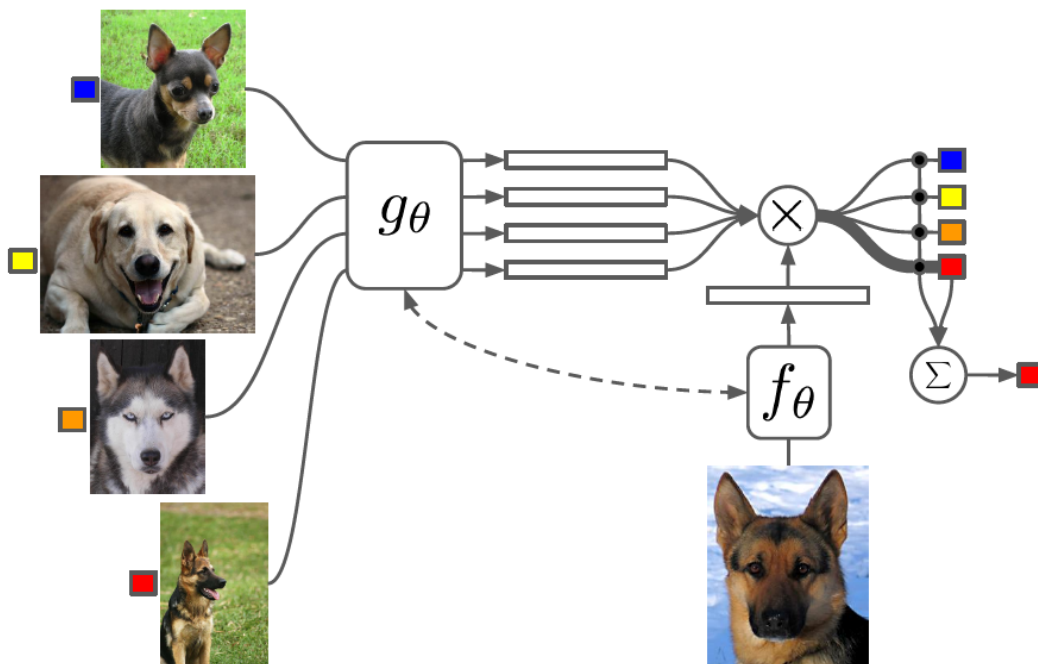


Figura 2.7: Esquemático das redes *Matching* usando uma configuração *4-way* e *1-shot*, com 1 imagem de consulta  $q$ .  
 Fonte: (VINNYALS et al., 2016)

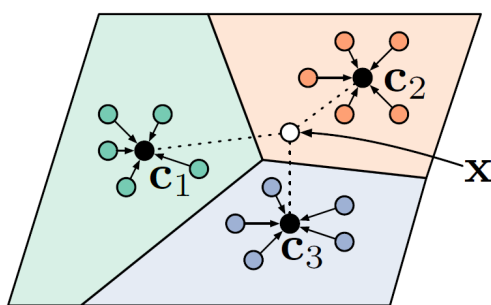


Figura 2.8: Estratégia das redes Prototípicas.  
 Fonte: (SNELL; SWERSKY; ZEMEL, 2017)

Embora as melhorias na extração de recursos de imagens tenham fornecido resultados promissores, a otimização dos modelos nem sempre é a principal motivação. Além disso, é importante que essas melhorias na arquitetura venham acompanhadas com melhorias no processo de medição de similaridade. O uso de outras estratégias de similaridade ainda não investigadas em FSL, por exemplo, pode contribuir com o desenvolvimento de novas tecnologias que lidam melhor com o paradigma de se reduzir a diferença entre imagens intraclasse e ampliar a diferença entre imagens interclasse, além de nortear pesquisadores em futuras pesquisas.

Apesar da grande relevância da abordagem métrica nos modelos de FSL, poucos estudos investigaram os efeitos de diferentes métricas de similaridade. Além disso, poucas métricas de similaridade foram de fato investigadas na maioria dos estudos. Yang et al. (2020), por exemplo, apresentaram um estudo para avaliação do desempenho das métricas  $L_1$ ,  $L_2$  e a distância do cosseno usando três modelos clássicos de FSL baseados em métrica, as *Matching networks*,

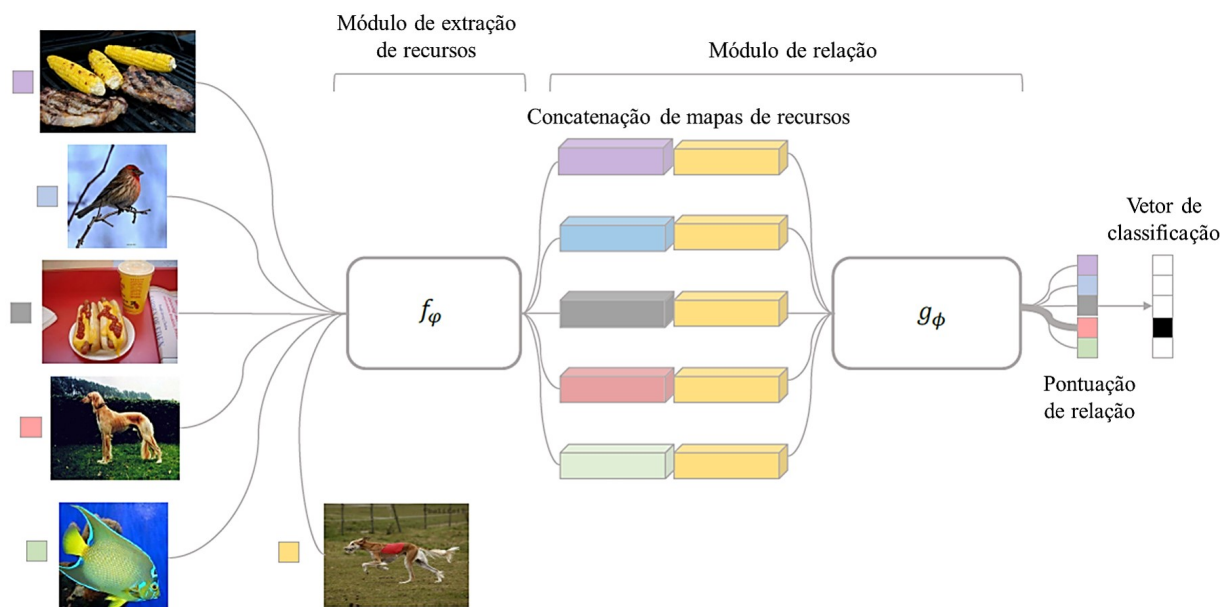


Figura 2.9: Rede de Relação demonstrada a partir de uma configuração 5-way e 1-shot, com 1 imagem de consulta  $q$ .

Fonte: Adaptado de (SUNG et al., 2018)

*Prototypical networks e Relation networks.* Seus resultados mostraram que as redes Prototípicas, com a métrica de similaridade  $L_2$  produzem melhores resultados gerais.

Dado que a maioria dos trabalhos publicados lida com métricas recorrentes na literatura, como a distância Euclidiana quadrática (SNELL; SWERSKY; ZEMEL, 2017; LI; YANG, 2020), a diferença absoluta (KOCH et al., 2015; YANG et al., 2020), e a similaridade do cosseno (VINYALS et al., 2016; CHEN et al., 2021), novos métodos devem ser investigados. Recentemente, os resultados dessas métricas foram melhorados combinando um extrator de recursos com um classificador baseado na distância de Mahalanobis regularizada (BATENI et al., 2020; NUTHALAPATI; TUNGA, 2021).

Melhorar a extração de recursos e a medição de similaridade é um caminho promissor para o avanço das técnicas de FSL. Este trabalho contribui com as pesquisas anteriores fornecendo um modelo com duas melhorias principais: 1) um estudo amplo sobre as divergências de Bregman em FSL para a adoção de uma abordagem relevante para medição de similaridade no modelo proposto (Capítulo 3), e 2) um módulo CNN de extração e fusão de recursos multicamadas para geração de representantes visuais consistentes (Capítulo 4).



# 3 ANÁLISE DE DIVERGÊNCIAS EM APRENDIZAGEM *FEW-SHOT*

Este capítulo trata de aspectos importantes para o aprendizado de algoritmos baseados em poucos dados, incluindo análise de divergências e de parâmetros *N-way* e *K-shot* dos modelos de *few-shot*. Uma arquitetura baseada nas redes Prototípicas foi construída para análise de classificação de insetos usando quatro divergências de Bregman. O conjunto de dados de insetos da agricultura IP-FSL, com categorias separadas por ciclo de maturidade do inseto, foi proposto para treinamento do modelo. O objetivo deste capítulo, é demonstrar caminhos promissores que minimizam os erros de classificação e fornecer um embasamento para a escolha da divergência usada como medida de similaridade em *few-shot*.

## 3.1 DIVERGÊNCIAS DE BREGMAN

As divergências de Bregman têm sido aplicadas com frequência em problemas de otimização, clusterização e aprendizagem de máquinas (SIAHKAMARI et al., 2020; BRÉCHETEAU; FISCHER; LEVRARD, 2021; CILINGIR; MANZELLI; KULIS, 2020). Esse grupo de divergências corresponde a uma abordagem que estabelece uma medida generalizada entre distribuições, definida nos termos de uma função estritamente convexa (BANERJEE et al., 2005).

**Definição:** Dada uma função continuamente diferenciável, estritamente convexa,  $F : S \rightarrow \mathbb{R}$ , definida em um domínio convexo  $S \subseteq \mathbb{R}^d$ , a divergência de Bregman entre  $\mathbf{x}, \mathbf{y} \in S$  induzida por  $F$ , é definida conforme a Equação 3.1.

$$D_F(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}) - F(\mathbf{y}) - \langle \mathbf{x} - \mathbf{y}, \nabla F(\mathbf{y}) \rangle \quad (3.1)$$

Em que  $\langle \mathbf{a}, \mathbf{b} \rangle$  indica o produto interno, e  $\nabla F(\mathbf{y})$  representa o vetor de gradiente de  $F$  avaliado em  $\mathbf{y}$ .

As divergências de Bregman possuem diversas propriedades fundamentais, dentre elas a não negatividade, isto é,  $D_F(\mathbf{x}, \mathbf{y}) \geq 0$ , em que  $D_F(\mathbf{x}, \mathbf{y}) = 0$  se e somente se  $\mathbf{x} = \mathbf{y}$ . Além disso, mesmo com algumas exceções, as divergências de Bregman são consideradas assimétricas, haja vista que  $D_F(\mathbf{x}, \mathbf{y}) \neq D_F(\mathbf{y}, \mathbf{x})$ . Nesse trabalho, avaliamos quatro divergências de Bregman no contexto de aprendizagem *few-shot*, amplamente utilizadas na literatura, as quais são descritas brevemente nesta seção.

### 3.1.1 Distância Euclidiana quadrática

Dentre as divergências de Bregman, a distância Euclidiana quadrática, gerada pela função estritamente convexa  $F(\mathbf{x}) = \|\mathbf{x}\|^2$ , é a mais simples e amplamente utilizada para a diferenciação de distribuições. Essa abordagem calcula a distância entre vetores de ponto a ponto. Dada a sua função convexa, a distância Euclidiana entre dois vetores  $\mathbf{x}$  e  $\mathbf{y}$  pode ser calculada conforme a Equação 3.2:

$$D_F(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2 \quad (3.2)$$

### 3.1.2 Distância quadrática de Mahalanobis

A distância de Mahalanobis, gerada pela função convexa  $F(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ , é definida como a medida de distância entre um ponto e uma distribuição. Por esse motivo, considera a covariância entre as variáveis envolvidas. A distância de Mahalanobis entre um vetor  $\mathbf{x}$  e uma distribuição  $\mathbf{y}$  pode ser calculada pela Equação 3.3, a qual é chamada distância de Mahalanobis quando  $\mathbf{A}$  é o inverso da matriz de covariância.

$$D_F(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \mathbf{A} (\mathbf{x} - \mathbf{y}) \quad (3.3)$$

A Equação 3.3 objetiva resolver o problema da distância Euclidiana quando os dados possuem correlação linear. Ela tem o efeito de transformar as variáveis em não correlacionadas, escalonando-as a partir da divisão pela matriz de covariância. Dessa forma, a equação 3.3 corresponde ao cálculo da distância Euclidiana com dados escalonados.

Em nosso trabalho, priorizamos o custo temporal para a estimativa da matriz de covariância. Por esse motivo, adotamos a hipótese de que a covariância baseada em protótipos de tarefa em vez de instâncias de classes, produz resultados significativos com baixo consumo de tempo durante o cálculo computacional. Portanto,  $\mathbf{x}$  corresponde ao conjunto de *embeddings* de consulta e  $\mathbf{y}$  o conjunto de protótipos de uma tarefa. Essa abordagem nos permite usar  $K \geq 1$  sem mudanças de paradigma no algoritmo de estimativa da covariância.

### 3.1.3 Divergência de Kullback-Leibler

A divergência de Kullback–Leibler (KL), ou entropia relativa, é gerada pela função convexa da entropia negativa  $\sum_{j=1}^d x_j \log_2 x_j$  para distribuições discretas. Trata-se de uma medida que quantifica a diferença entre duas distribuições de probabilidade. A divergência de Bregman entre duas distribuições de probabilidade discretas, correspondente a função convexa geradora da divergência KL é descrita pela Equação 3.4.

$$D_F(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^d x_j \log_2\left(\frac{x_j}{y_j}\right) = KL(\mathbf{x}||\mathbf{y}) \quad (3.4)$$

Em nossos experimentos, transformamos cada *embedding* ou protótipo de classe em uma espécie de vetor de probabilidade para o cálculo da divergência KL. Em outras palavras, tomamos um vetor de características  $\mathbf{y}$  e o transformamos em um vetor de probabilidade calculado conforme a Equação 3.5, de forma que  $\sum_{j=1}^d y_j = 1$ .

$$\mathbf{y} = \frac{\mathbf{y}'}{\sum_{j=1}^d y'_j} \quad (3.5)$$

Em que  $\mathbf{y}'$  é o vetor inicial.

Observa-se que pode existir, na Equação 3.4, um problema relacionado ao cálculo do logaritmo quando houver valores iguais à zero. Por esse motivo, uma constante  $\epsilon$  com um pequeno valor foi adicionada aos vetores antes do cálculo da divergência, para evitar resultado infinito negativo ( $\log 0$ ), ou mesmo a divisão por zero.

### 3.1.4 Divergência de Itakura-saito

A divergência de Itakura-Saito (IS) é gerada pela função do logaritmo negativo  $F(x) = -\log x$ . Assim como a divergência KL, IS é uma medida assimétrica e amplamente utilizada em processamento de sinais. A divergência IS é definida conforme a Equação 3.6.

$$D_F(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^d \frac{x_j}{y_j} - \log\left(\frac{x_j}{y_j}\right) - 1 = IS(\mathbf{x}||\mathbf{y}) \quad (3.6)$$

Assim como no cálculo da divergência KL, foram gerados vetores de probabilidade, conforme a Equação 3.5, para cálculo da divergência IS nas tarefas de classificação.

Esta seção mostrou as principais divergências de Bregman utilizadas para comparação de distribuições e como foram utilizadas neste trabalho. A próxima seção trata do modelo de *few-shot* adotado para estudo dessas divergências.

## 3.2 O MODELO ADOTADO PARA ANÁLISE DE DIVERGÊNCIAS

A rede Prototípica (ProtoNet) (SNELL; SWERSKY; ZEMEL, 2017) é um dos algoritmos baseados em métrica mais utilizados e bem sucedidos na literatura. Adotamos um modelo de aprendizagem baseado nessa arquitetura para investigação das divergências de Bregman em FSL.

A arquitetura da ProtoNet é constituída de uma CNN para extração de características de imagens e um comparador métrico baseado na distância Euclidiana. A CNN aprende características de imagens a partir da métrica de similaridade por distância para a previsão de uma ou mais classes. Embora as redes prototípicas usem originalmente uma divergência de Bregman (distância Euclidiana) como medida de similaridade, sua estrutura permite a utilização de outras métricas capazes de diferenciar vetores, dando suporte a ampla utilização das Divergências de Bregman.

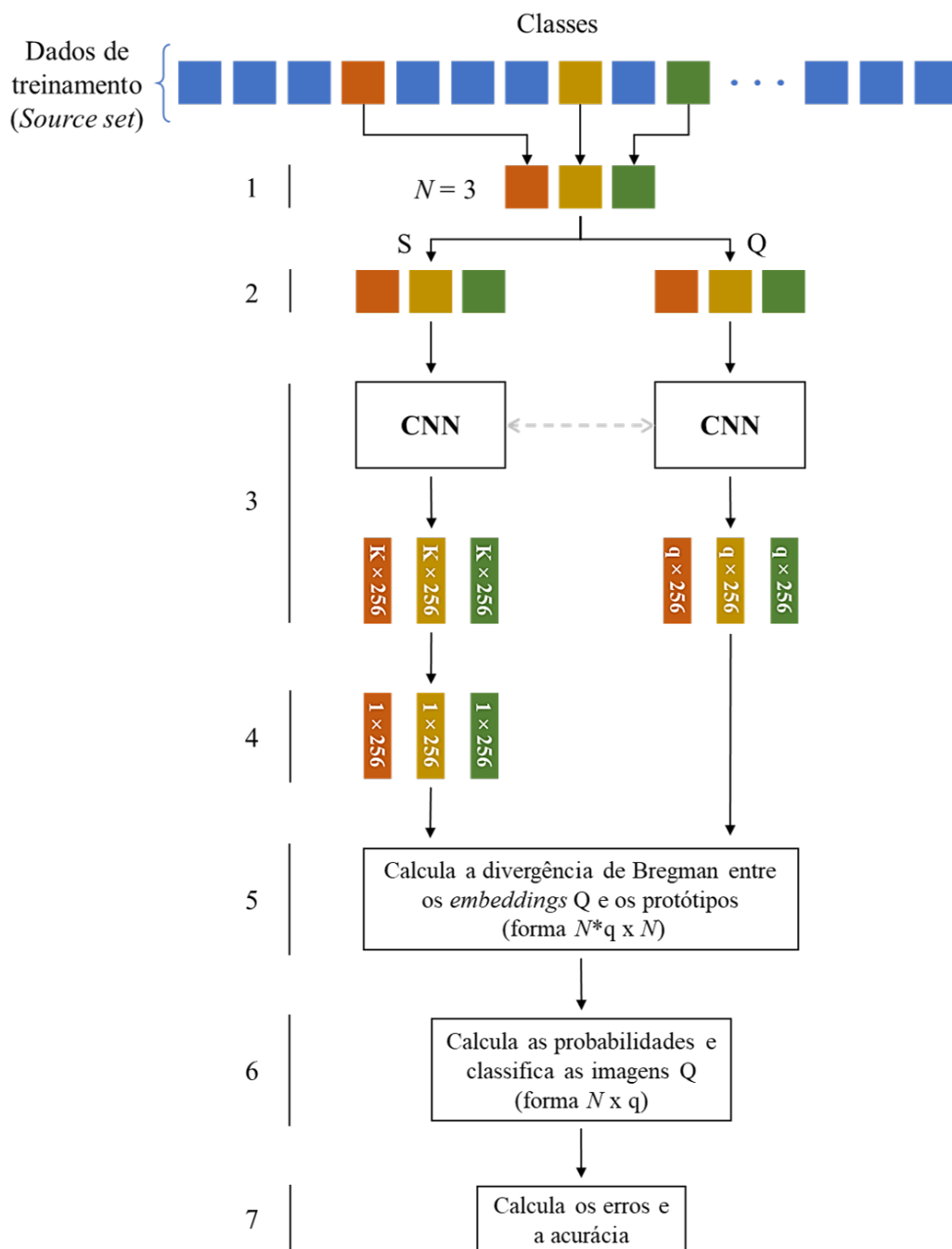


Figura 3.1: Episódio de treinamento.

Modificamos a estrutura original das redes prototípicas para obtenção de um aprendizado eficiente, com melhor poder de extração de características de imagens e redução de superajustes aos

dados de treinamento. Cada um dos quatro blocos convolucionais contém uma camada convolucional de 64 filtros  $3 \times 3$ , camada de normalização, ReLU e max-pooling  $3 \times 3$ . A saída do modelo compreende a camada *Flatten*, que resulta em um vetor de 256 dimensões.

O objetivo do modelo apresentado nesse capítulo é aprender características de imagens para diferenciar classes por meio das divergências de Bregman. Além da distância Euclidiana, investigamos os resultados de outras três divergências: a distância de Mahalanobis e as divergências de Kullback-Leibler e Itakura-Saito. O nosso modelo integra, portanto, essas divergências como medidas de dissimilaridade para a classificação de insetos dos dados IP-FSL.

As diretrizes da meta-aprendizagem foram usadas para a condução dos experimentos nesse trabalho. As etapas de treinamento do modelo são inicializadas selecionando-se, aleatoriamente,  $N$  classes do conjunto de origem para treinamento de uma tarefa de classificação. A Figura 3.1 mostra o esquema de uma tarefa, a qual representa um episódio de treinamento do nosso modelo. A seguir, explicamos cronologicamente o fluxo de treinamento com índices indicando cada etapa da Figura 3.1:

1. *Seleção aleatória das classes de uma tarefa de classificação*: Inicialmente,  $N$  classes são selecionadas aleatoriamente do conjunto de dados (*source set*) para treinamento de uma tarefa  $N$ -way- $K$ -shot. A Figura 3.1 representa uma tarefa 3-way.
2. *Divisão dos dados em conjuntos de suporte e consulta*: Os dados da respectiva tarefa são divididos em conjunto de suporte  $S$  e de consulta  $Q$ , conforme os valores dos parâmetros  $N$ ,  $K$  e  $q$  previamente atribuídos. Nessa etapa, os índices das classes (0, 1, ...,  $N$ -way - 1) são definidos para a criação das variáveis *Target* usadas para a verificação do erro de classificação.
3. *Geração de vetores de recursos de imagens, os Embeddings*: As características de todas as imagens de uma tarefa são extraídas pela CNN, formando os *embeddings* de suporte e de consulta.
4. *Geração de Protótipos*: Os protótipos são calculados a partir dos *embeddings* de suporte e funcionam como representantes de classes. Cada protótipo corresponde ao centroide dos *embeddings* de uma classe, calculado conforme a Equação 3.7.

$$c^{(n)} = \frac{1}{K^{(n)}} \sum_{i=1}^{K^{(n)}} f_{\Phi}(\mathbf{x}_i^{(n)}) \quad (3.7)$$

Em que  $c^{(n)}$  representa o centroide da classe  $n$ .

5. *Cálculo das divergências*: Nessa etapa, a divergência de Bregman é calculada para medição de similaridade entre os *embedding* de consulta e os protótipos, conforme o caso adotado: distância Euclidiana, distância de Mahalanobis, divergência de Kullback–Leibler ou a divergência de Itakura–Saito.

6. *Calculo das probabilidades e classificação de imagens*: As imagens de consulta são rotuladas de acordo com uma distribuição de probabilidade sobre as divergências. Uma menor distância indica maior probabilidade de uma imagem ‘q’ pertencer a uma determinada classe. A probabilidade é obtida por meio do *softmax*, conforme a Equação 3.8.

$$p_{\phi}(y = n|\mathbf{x}) = \frac{\exp(-d(f_{\Phi}(\mathbf{x}), c^{(n)}))}{\sum_{n'} \exp(-d(f_{\Phi}(\mathbf{x}), c^{(n')}))} \quad (3.8)$$

7. *Perda e retropropagação*: O aprendizado das redes Prototípicas é realizado minimizando a probabilidade de log negativo  $J(\phi) = -\log_{\phi}(y = n|\mathbf{x})$  por meio do algoritmo de gradiente descendente estocástico (SGD).

As etapas acima descreveram sucintamente o treinamento de um episódio do modelo aqui utilizado. O modelo completamente treinado é obtido após diversas tarefas de classificação, cujos detalhes são apresentados na seção 3.4. Um episódio de teste é semelhante ao de treinamento, com exceção de que os parâmetros do modelo são congelados durante o teste. Além disso, o teste é realizado usando o *target set*. Por fim, o episódio termina com a classificação das imagens e a obtenção da acurácia com respeito ao conjunto de consulta Q da tarefa de teste. Aqui a acurácia é obtida pela média das acurácias em várias tarefas de teste.

### 3.3 O CONJUNTO DE DADOS IP-FSL

Atualmente, existem poucos conjuntos de dados de insetos disponíveis na literatura. Em se tratando de conjuntos de dados específicos para o aprendizado *few-shot* no meio agrícola, o número pode ser considerado irrisório. Recentemente, Wu et al. (2019) publicaram os dados IP102 contendo 102 classes de insetos divididas entre 8 culturas (*rice, corn, wheat, beet, alfafa, vitis, citrus, mango*), com o total de 75222 imagens. O IP102 possui imagens com condições de captura diversas; algumas de cenas reais, como as capturadas em campo (insetos na planta, com fundo de imagem variado), outras em condições controladas como em laboratório, bem como imagens contendo desenhos de insetos. Além disso, os dados IP102 possuem imagens diversificadas em relação à maturidade dos insetos. Em outras palavras, algumas classes possuem imagens do inseto adulto, bem como outras imagens de suas larvas, o que pode dificultar a busca de padrões de uma determinada classe, e como consequência, não permitir um aprendizado adequado ao algoritmo de aprendizagem de máquinas.

A medida que novos conjuntos de dados são disponibilizados na literatura, há um ganho no desenvolvimento de novas metodologias para aplicação na agricultura. Por esse motivo, analisamos exaustivamente todo o conjunto de dados IP102 para a construção de um novo conjunto de dados de insetos (IP-FSL), específico para a aprendizagem *Few-shot* e com atenção à fase biológica da praga. Acreditamos que o treinamento do modelo direcionado à fase biológica do inseto garante um aprendizado otimizado, com maior ganho em relação à acurácia e em temos

práticos, principalmente no que diz respeito ao reconhecimento precoce da praga na fase inicial de vida, contribuindo para a contenção do avanço das pragas nas lavouras.

O conjunto de dados IP102 possui classes contendo insetos em diferentes estágios de maturidade. Para a construção do novo conjunto de dados com distinção sobre o ciclo de vida do inseto, selecionamos imagens para formar novas classes atribuindo-as em duas categorias: a fase primária e a fase adulta. Considerou-se da fase primária aquelas imagens que contêm a presença do ovo, larva, lagarta ou da pupa. Já a fase adulta inclui insetos jovens e adultos. Para aquelas espécies com imagens das duas fases citadas, foram formadas novas classes. As imagens foram selecionadas da seguinte maneira: classes contendo imagens de insetos de uma única fase permaneceram as mesmas dentro de sua categoria; já as classes que possuem as duas fases do inseto foram divididas para atribuição das imagens, sendo alocadas, portanto, nas respectivas categorias. Em outras palavras, uma determinada espécie de inseto que contenha imagens de ambas as fases, foram atribuídas às categorias primária e adulta.

Decidiu-se adotar um número mínimo de 15 e máximo de 50 imagens para compor cada classe do conjunto de dados IP-FSL. Tais quantidades são suficientes para treinamento de algoritmos de *few-shot*. Como critérios de seleção, considerou-se, principalmente, as imagens com condições de captura em campo, com prioridade daquelas que possuem insetos com as dimensões muito pequenas em relação às dimensões das fotografias. Devido à quantidade pequena de imagens disponíveis em algumas classes do conjunto de dados fonte, que corresponderam aos critérios de seleção adotados, algumas classes não alcançaram o montante de 50 imagens.

A configuração final dos dados IP-FSL, representada na Figura 3.2, conta com um total de 6817 imagens. O subconjunto 'Fase primária' contém 45 classes de insetos no estágio inicial de vida, totalizando 2050 imagens. Já o subconjunto 'Fase adulta' é composto de 97 classes de insetos, totalizando 4767 imagens. A Figura 3.3 mostra alguns exemplos dos dados IP-FSL. As informações de cada classe e a quantidade de imagens são apresentadas na Tabela A.1.

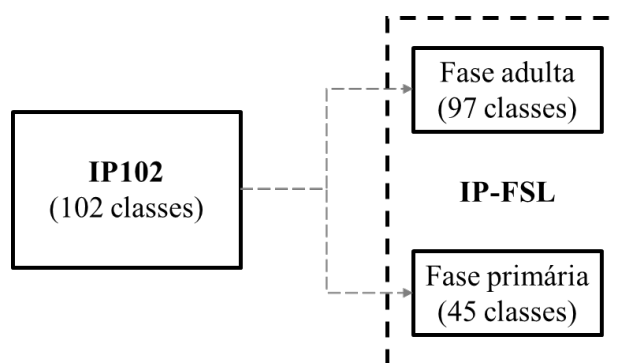


Figura 3.2: Configuração dos dados IP-FSL.

O IP-FSL é um novo conjunto de dados de insetos, construído especificamente para aprendizagem *few-shot* (GOMES; BORGES, 2022). No entanto, os dados podem ser usados para a aprendizagem de outros modelos clássicos em aprendizagem de máquinas. Para os experimentos deste capítulo, as classes dos dados da fase adulta foram divididas em uma taxa de 80:20 para

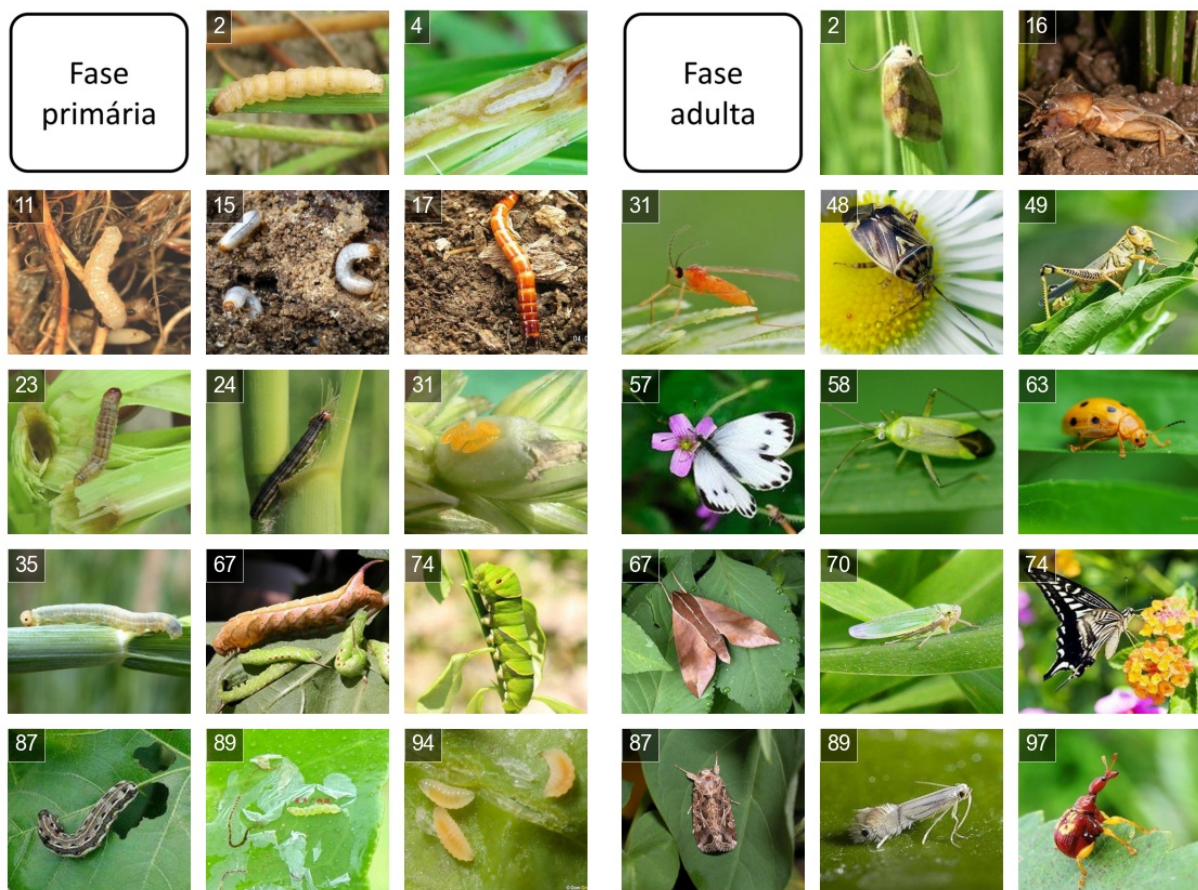


Figura 3.3: Algumas amostras dos dados IP-FSL. Os rótulos de classes estão numerados conforme os nomes dos insetos na Tabela A.1. A figura mostra alguns exemplos de classes que possuem insetos das duas fases, por isso aparecem nos dois subconjuntos (são eles, 2, 31, 67, 74, 87 e 89).

a formação dos conjuntos *Source* e *Target* respectivamente. Os dados da fase primária foram divididos sob uma taxa aproximada de 77:23 para treinamento e teste do modelo nessa categoria.

### 3.4 EXPERIMENTAL

Todas as configurações de treinamento e de teste descritas nesta seção foram igualmente realizadas para os experimentos usando os dois subconjuntos de dados contidos no IP-FSL. O modelo de aprendizagem recebe imagens com três canais (RGB), sem qualquer processamento de imagem. Realizou-se, entretanto, algumas transformações morfológicas para padronização e aumento do número de classes.

Inicialmente, todas as imagens do conjunto de dados foram redimensionadas para o formato  $96 \times 96 \times 3$ . Conforme sugerido no artigo original das redes Prototípicas, para melhorias no desempenho do modelo, as imagens foram rotacionadas em múltiplos de  $90^\circ$  para formação de novas classes de insetos. Dessa forma, após múltiplas rotações até  $270^\circ$ , cada subconjunto de dados passou a ter quatro vezes o número inicial de classes, não interferindo no quantitativo de



instâncias dentro das respectivas classes. Ao ser inserida uma imagem redimensionada à arquitetura de extração de característica do modelo, um vetor de dimensão 256 é retornado.

Os experimentos foram realizados para avaliação dos resultados sob dois aspectos: a variação dos parâmetros  $N$ -way- $K$ -shot e os efeitos das Divergências de Bregman na tarefa de classificação de insetos. Para a investigação dos efeitos na variação dos parâmetros  $N$ -way,  $K$ -shot, adotou-se uma abordagem conforme utilizada por Li e Yang (2021) em outro contexto. Para todas as divergências de Bregman usadas nessa pesquisa, portanto, foram analisadas as configurações  $1$ -shot,  $5$ -shot e  $10$ -shot em tarefas de  $3$ -way e  $5$ -way. Os valores  $N$ -way foram analisados permutando-os de forma que  $N_{train}, N_{test} = 3, 5$  e vice-versa. Além disso,  $q = 5$  permaneceu fixo para todos os experimentos.

Os experimentos foram conduzidos por meio da plataforma Google Colaboratory. A biblioteca Pytorch na versão 1.9.0 foi usada para treinamento dos modelos. Os dados de origem (*source set*) foram usados para treinamento de cada modelo com a sua respectiva divergência por 10 épocas, sendo 2000 episódios por época em cada uma das duas configurações  $N$ -way. A taxa de aprendizagem inicial de  $10^{-3}$  decaiu pela metade a cada época. Os testes foram realizados usando os dados de destino (*target set*).

### 3.5 RESULTADOS E DISCUSSÕES

Uma extensa cadeia experimental foi estruturada para abranger a avaliação das divergências de Bregman e dos parâmetros  $N$ -way e  $K$ -shot em FSL. Os experimentos foram realizados usando dados de insetos minuciosamente selecionados, para simular com maior proximidade as situações reais de campo. Além disso, usando a metodologia de meta-aprendizagem, os modelos foram treinados usando os dados de origem (*source*) e testados com os dados de destino (*target*), de forma que não houve nenhuma relação de classes entre ambos os dados. Em aprendizagem *Few-shot* os resultados são comumente apresentados como a média das classificações de diversas tarefas de teste. Os resultados apresentados aqui correspondem à acurácia média de 1000 episódios de teste.

Para cada uma das quatro divergências de Bregman investigadas nesse trabalho, foram realizadas 12 categorias de experimentos. Em resumo, foram realizados 48 experimentos em cada um dos dois subconjuntos dos dados IP-FSL. As Tabelas 3.1 e 3.2 apresentam as configurações e resultados experimentais usando os dados de insetos na fase adulta e na fase primária respectivamente. Tais tabelas mostram os resultados das divergências de Bregman separados em colunas com respeito às variações dos parâmetros  $N$ -way- $K$ -shot. A partir desses dados, analisamos a influência das divergências e da variação dos parâmetros  $N$ -way e  $K$ -shot em todos os experimentos.

Os resultados desse capítulo foram analisados sob os seguintes aspectos: inicialmente, duas seções foram dedicadas à análise dos efeitos da variação dos parâmetros  $N$ -way- $K$ -shot, a partir dos gráficos nas Figuras 3.4 e 3.5; em seguida, os resultados das divergências de Bregman foram comparados e discutidos a partir das Figuras 3.7 e 3.8.

Tabela 3.1: Resultados da classificação de insetos na fase adulta.

$N_{train}, N_{test}$	Euclidiana			Mahalanobis			KL			IS		
	$K=1$	$K=5$	$K=10$	$K=2$	$K=5$	$K=10$	$K=1$	$K=5$	$K=10$	$K=1$	$K=5$	$K=10$
3, 3	0,7434	0,8435	0,8682	0,7568	0,8527	0,8667	0,7797	0,8633	0,8821	0,7595	0,8471	0,8717
3, 5	0,6013	0,7510	0,7782	0,6144	0,7409	0,7645	0,6252	0,7610	0,7899	0,6144	0,7552	0,7943
5, 3	0,7704	0,8520	<b>0,8806</b>	0,7607	0,8555	<b>0,8731</b>	0,7831	0,8722	<b>0,8839</b>	0,7813	0,8729	<b>0,8803</b>
5, 5	0,6216	0,7615	0,7958	0,6321	0,7476	0,7782	0,6694	0,7743	0,8099	0,6580	0,7768	0,8083

Tabela 3.2: Resultados da classificação de insetos na fase primária.

$N_{train}, N_{test}$	Euclidiana			Mahalanobis			KL			IS		
	$K=1$	$K=5$	$K=10$	$K=2$	$K=5$	$K=10$	$K=1$	$K=5$	$K=10$	$K=1$	$K=5$	$K=10$
3, 3	0,8045	0,8621	0,8813	0,7920	0,8670	<b>0,8905</b>	0,8167	0,8791	<b>0,9010</b>	0,8107	0,8778	0,8869
3, 5	0,6723	0,7819	0,8214	0,6595	0,7687	0,8025	0,6797	0,7835	0,8132	0,6812	0,7840	0,7989
5, 3	0,7940	0,8654	<b>0,8903</b>	0,7950	0,8756	0,8865	0,8133	0,8779	<b>0,9010</b>	0,8018	0,8798	<b>0,8976</b>
5, 5	0,6758	0,7722	0,8112	0,6786	0,7780	0,7925	0,6906	0,8072	0,8162	0,6859	0,8044	0,8248

### 3.5.1 Os efeitos da variação de $K$ -shot

O parâmetro  $K$ -shot representa a quantidade de imagens de suporte em cada classe dentro de uma tarefa de classificação. Como mostram os gráficos das Figuras 3.4 e 3.5, a acurácia possui correlação positiva em relação a  $K$ -shot, independentemente do tipo de dado (fase adulta ou fase primária), visto que as tarefas de  $1$ -shot produziram as menores acurácias, sendo gradativamente superadas pelas tarefas  $5$ -shot e  $10$ -shot.

Objetivamente, o modelo visa armazenar nos protótipos a maior quantidade de informação possível sobre as respectivas classes. Naturalmente, quanto maior o valor de  $K$ , mais informações sobre as distribuições de classes serão captadas pelos protótipos, aumentando a sua representatividade. A razão disso, é que mais informações são aprendidas pela rede CNN a medida que o número de imagens exploradas aumenta. Esse é um típico cenário da aprendizagem clássica, em que o aprendizado do modelo tende a aumentar a medida que o número de instâncias de classes aumenta. No contexto de aprendizagem  $few$ -shot, é importante que o número de instâncias  $K$  de uma classe não seja elevado, normalmente esse parâmetro varia entre 1 e 10.

### 3.5.2 Os efeitos da variação de $N$ -way

O parâmetro  $N$ -way representa a quantidade de classes dentro de uma tarefa de classificação, para as quais o modelo precisa rotular as imagens de consulta. Conhecer o desempenho do modelo multiclasse com respeito à configuração  $N$ -way é fundamental. Os parâmetros  $N_{train}$  e  $N_{test}$  foram variados, assumindo os números 3 e 5, para a análise de quais circunstâncias o modelo tem maior robustez para classificação das imagens dos dados IP-FSL. Com relação aos resultados obtidos e apresentados por meio das Figuras 3.4 e 3.5, pode-se destacar três situações relevantes para análise:  $N_{train} < N_{test}$ ,  $N_{train} = N_{test}$  e  $N_{train} > N_{test}$ .

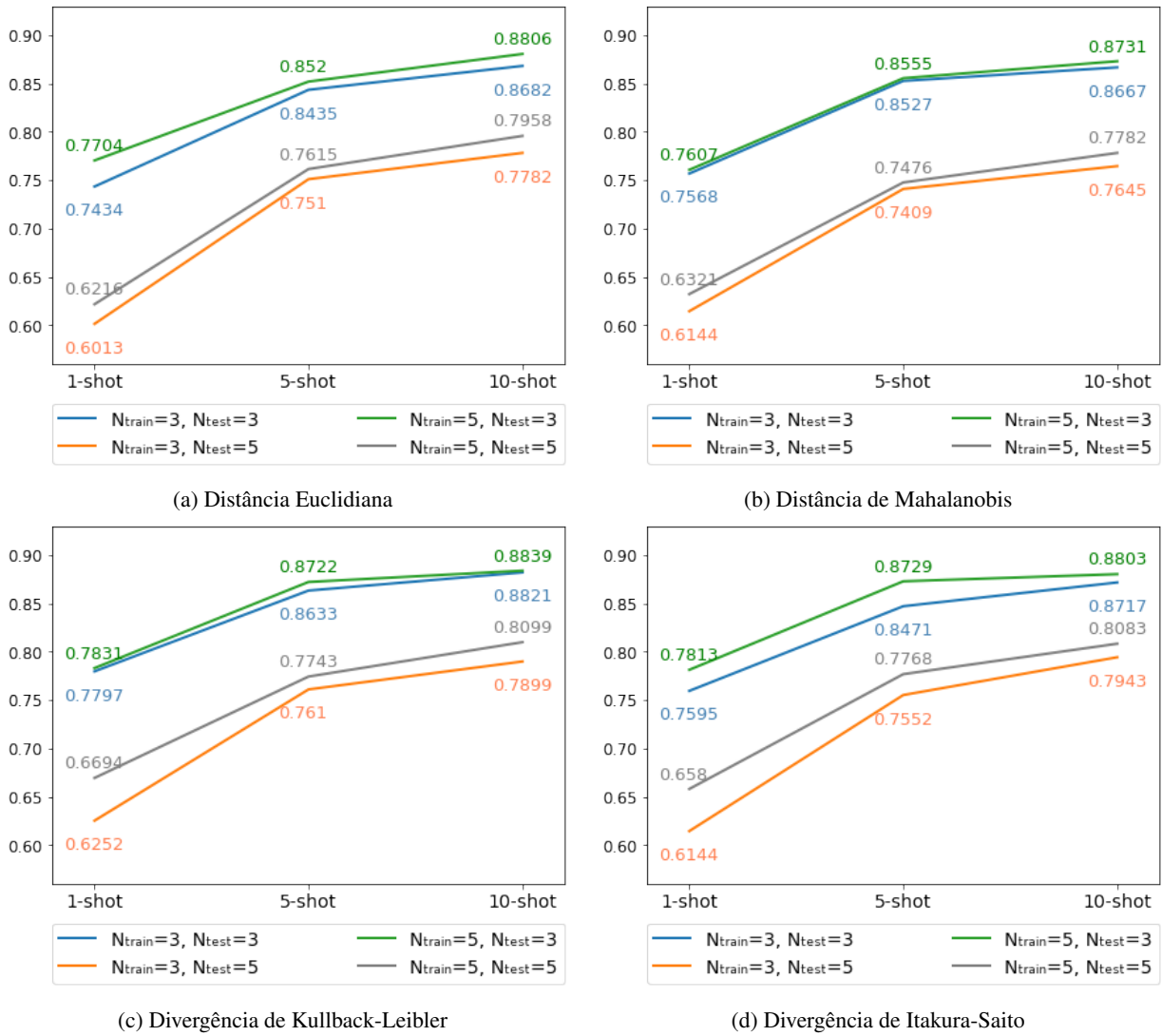


Figura 3.4: Tendência de acurácia em relação aos parâmetros  $N$ -way- $K$ -shot para classificação de insetos da fase adulta.

### 3.5.2.1 Situação 1 ( $N_{train} = N_{test}$ )

Notavelmente, os parâmetros configurados como  $N_{train} = N_{test} = 3$  produziram maior acurácia em comparação com a configuração  $N_{train} = N_{test} = 5$  em todas as divergências, conforme mostraram as Figuras 3.4 e 3.5. Isso significa que um menor número de classes dentro de uma tarefa produz menores erros de classificação. Dessa forma, pode-se dizer que a dificuldade de aprendizagem do modelo é proporcional a quantidade de classes existentes em uma tarefa, ou seja, quanto maior  $N_{train}$ , maior a dificuldade de aprendizagem.

Portanto, a dificuldade de aprendizagem do modelo, referida acima, pode ser quantificada com respeito ao montante de classes exploradas na tarefa de classificação, associado à semelhança existente entre algumas dessas classes. Como os modelos de *few-shot* são treinados e testados por meio de tarefas, essa definição que quantifica a dificuldade de aprendizado, e por consequência a dificuldade de classificação, é válida tanto para o treinamento quanto para o teste.

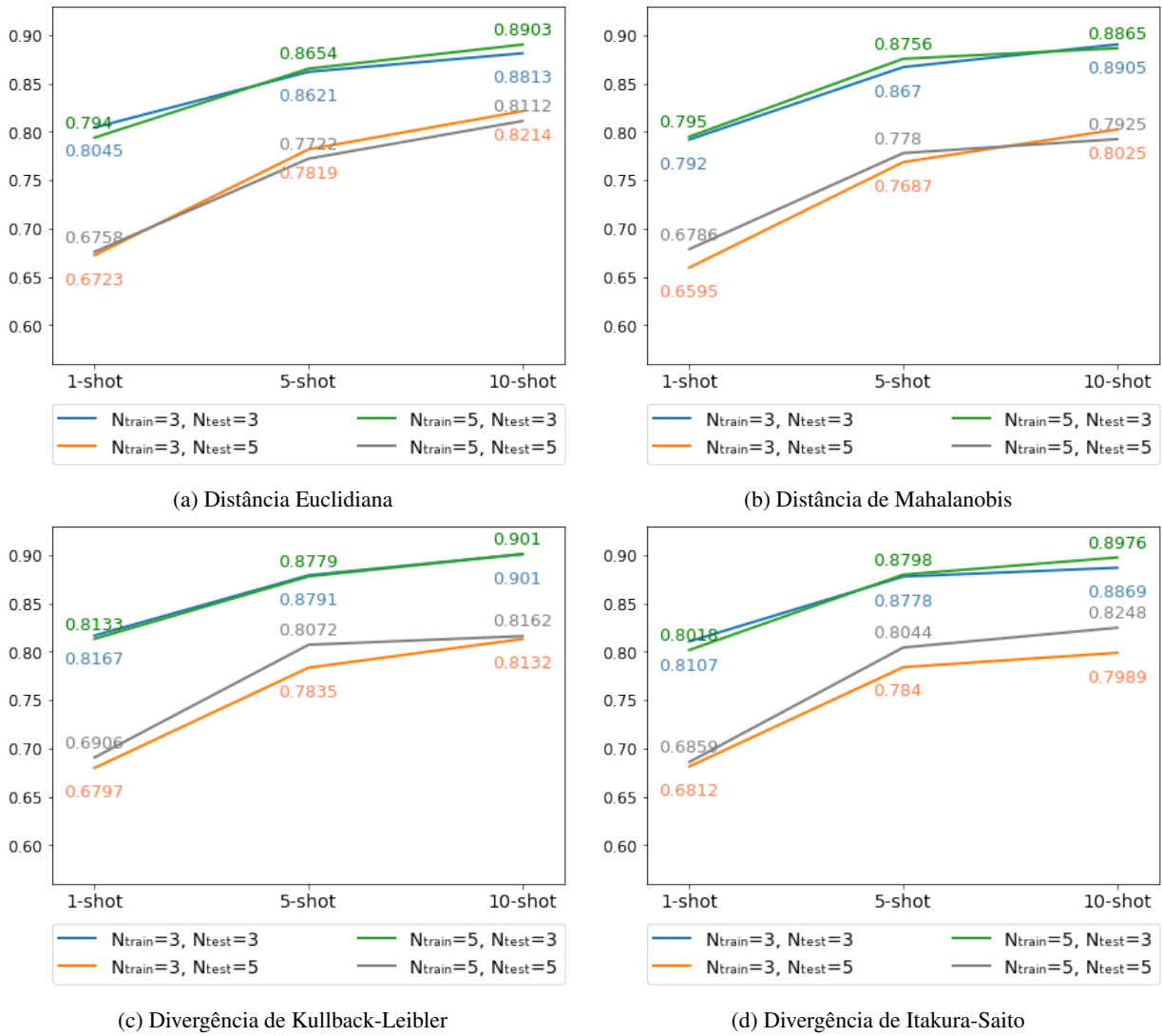


Figura 3.5: Tendência de acurácia em relação aos parâmetros  $N$ -way- $K$ -shot para classificação de insetos da fase primária.

### 3.5.2.2 Situação 2 ( $N_{train} < N_{test}$ )

Como apontaram os gráficos das Figuras 3.4 e 3.5, a acurácia cai consideravelmente no caso de  $N_{train} < N_{test}$ . Com referência à definição de dificuldade de aprendizagem citada acima, pode-se dizer que o modelo aprende em um ambiente de baixa dificuldade de aprendizagem, visto que realiza classificações entre uma variedade baixa de classes, ao passo que é testado em um ambiente de difícil classificação devido ao maior número de classes para escolha, por isso a acurácia tende a cair. Em outras palavras, o modelo não aprendeu o suficiente para classificar a partir de uma situação de maior dificuldade. A Figura 3.6 mostra o desempenho da aprendizagem do modelo usando a divergência KL. Nota-se que a curva de perda de treinamento é elevada quando  $N_{train} = 5$ , e conseqüentemente a acurácia é baixa. O contrário ocorre quando  $N_{train} = 3$ , em que a perda de treinamento é baixa e a acurácia de treinamento é maior.

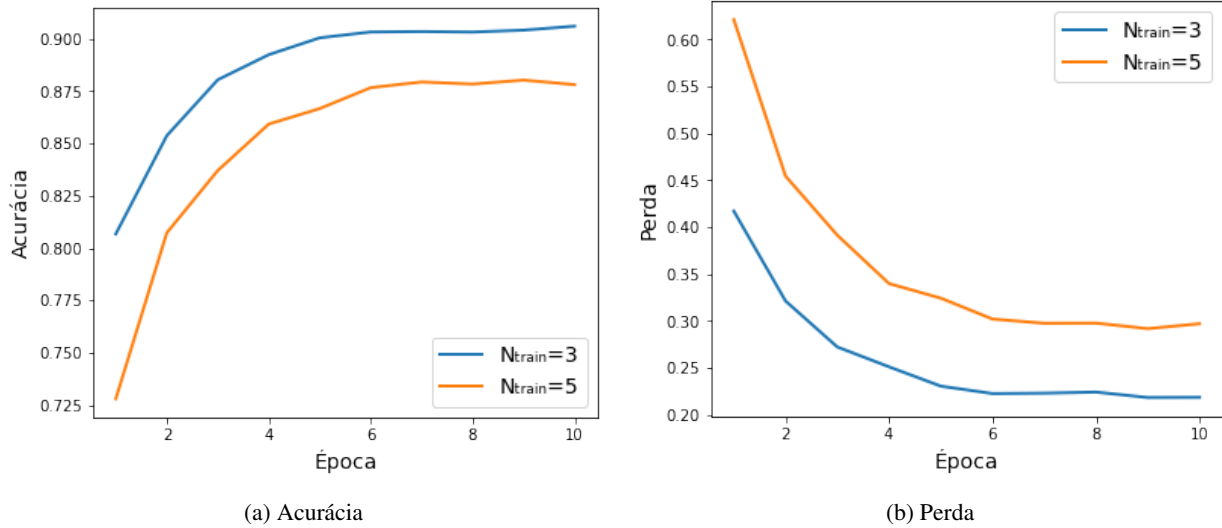


Figura 3.6: Um exemplo da perda e acurácia de treinamento usando o conjunto de dados da fase adulta e a divergência KL em uma configuração de 5-shot.

### 3.5.2.3 Situação 3 ( $N_{train} > N_{test}$ )

Tendo em vista os dados no decorrer do treinamento, conforme mostra a Figura 3.6, pode-se dizer que o modelo aprende com maior dificuldade quando  $N_{train} = 5$ , já que há um número maior de classes nas tarefas de classificação. Observa-se que há uma redução considerável na perda durante o treinamento após 10 épocas, entretanto, o mínimo é alcançado ainda em um patamar elevado. Por outro lado, na configuração  $N_{train} > N_{test}$  o modelo é testado em um ambiente com menor variedade de classes para a atribuição das imagens de consulta, aumentando a probabilidade de classificação correta. Dessa forma, a classificação usando os dados de teste terá menor dificuldade e conseqüentemente a acurácia será elevada.

### 3.5.2.4 Configurações importantes de $N$ -way para classificação de insetos

Em se tratando do conjunto de insetos adultos, os resultados da variação do parâmetro  $N$ -way sugerem que o treinamento usando  $N_{train} > N_{test}$  é mais vantajoso apesar de muito próximo dos resultados de  $N_{train} = N_{test} = 3$ . O primeiro caso citado superou o segundo em todos os testes, como mostra a Figura 3.4. Apesar disso, não é possível definir até que ponto  $N_{train} > N_{test}$  seja a melhor escolha visto que o escopo dos experimentos se limitou à variação  $N$ -way = 3 e 5. Entretanto, dentro dos valores estudados,  $N_{train} = 5$  em conjunto com  $N_{test} = 3$  é uma escolha razoável para os parâmetros do modelo de classificação *few-shot* de insetos adultos.

Considerando os resultados para os dados da fase primária (Figura 3.5), não foi possível confirmar qual a melhor configuração entre  $N_{train} > N_{test}$  e  $N_{train} = N_{test} = 3$ . Com exceção da divergência KL, todas as distâncias registraram inversões nos melhores resultados ao longo da variação de  $K$ -shot. Entretanto, as variações são muito pequenas, com resultados muito aproximados. Possivelmente, a semelhança entre algumas classes dentro da mesma tarefa, pode fazer

com que o desempenho oscile ligeiramente. As classes de lagartas, por exemplo, são muito semelhantes em alguns casos, e com isso, as divergências também podem ser aproximadas, o que dificulta a classificação correta.

### 3.5.3 Análise das divergências de Bregman

A adoção de uma abordagem métrica adequada para um modelo de classificação *few-shot* é um fator importante para a obtenção do melhor desempenho, tanto para os comparadores por pares quanto para os modelos de classificação multiclasse. Por esse motivo, estudar os efeitos das diversas metodologias de medição de similaridade entre dados é fundamental para a definição do equilíbrio entre o tempo de processamento no treinamento do modelo, a capacidade computacional disponível e acurácia desejada.

A partir dos diversos experimentos realizados nesse capítulo, foi possível avaliar o comportamento do modelo de FSL com respeito às quatro divergências de Bregman usadas como medidas de similaridade. Os resultados podem fornecer um direcionamento para a escolha da divergência adequada usada nos modelos *few-shot* tanto para classificação de insetos na agricultura quanto para outras aplicações envolvendo classificação visual. Nas Figuras 3.7 e 3.8 comparamos os resultados das divergências para os dados de insetos adultos e do estágio primário respectivamente. As próximas seções abordam as análises de cada divergência separadamente. Em seguida, uma análise comparativa de sensibilidade e desempenho relacionando o tempo de processamento e acurácia é realizada para maiores esclarecimentos.

#### 3.5.3.1 Distância Euclidiana

A distância Euclidiana foi uma métrica acertadamente escolhida pelos autores das redes Prototípicas. Dentre as características mais importantes dessa métrica, destaca-se a simplicidade do método para medição da distância entre pontos, assumindo o posto de mais simples e utilizada entre as divergências de Bregman. Nos experimentos aqui apresentados, a distância Euclidiana produziu resultados gerais ligeiramente inferiores aos das divergências KL e IS. Embora a distância Euclidiana tenha superado a IS em 0,1% em um ponto específico da classificação de insetos adultos,  $N_{train} = 5$ ,  $N_{test} = 3$  e *10-shot* (Figura 3.7c), e tenha superado todas as divergências usando a configuração  $N_{train} = 3$ ,  $N_{test} = 5$  e *10-shot* (Figura 3.8b) nos dados da fase primária, isso não ocorreu em nenhuma outra configuração.

Os resultados que definem a métrica mais vantajosa entre as distâncias Euclidiana e de Mahalanobis variaram sem um padrão de configuração *N-way-K-shot* para que isso ocorresse. Essas variações não permitiram a definição de que uma métrica possa ser mais viável, do ponto de vista da acurácia, em relação à outra. Provavelmente, os efeitos relacionados a medida de variabilidade entre as variáveis dentro de uma tarefa de classificação, expressados na matriz de covariância usada no cálculo da distância de Mahalanobis, fizeram com que a distância Euclidiana tenha sido

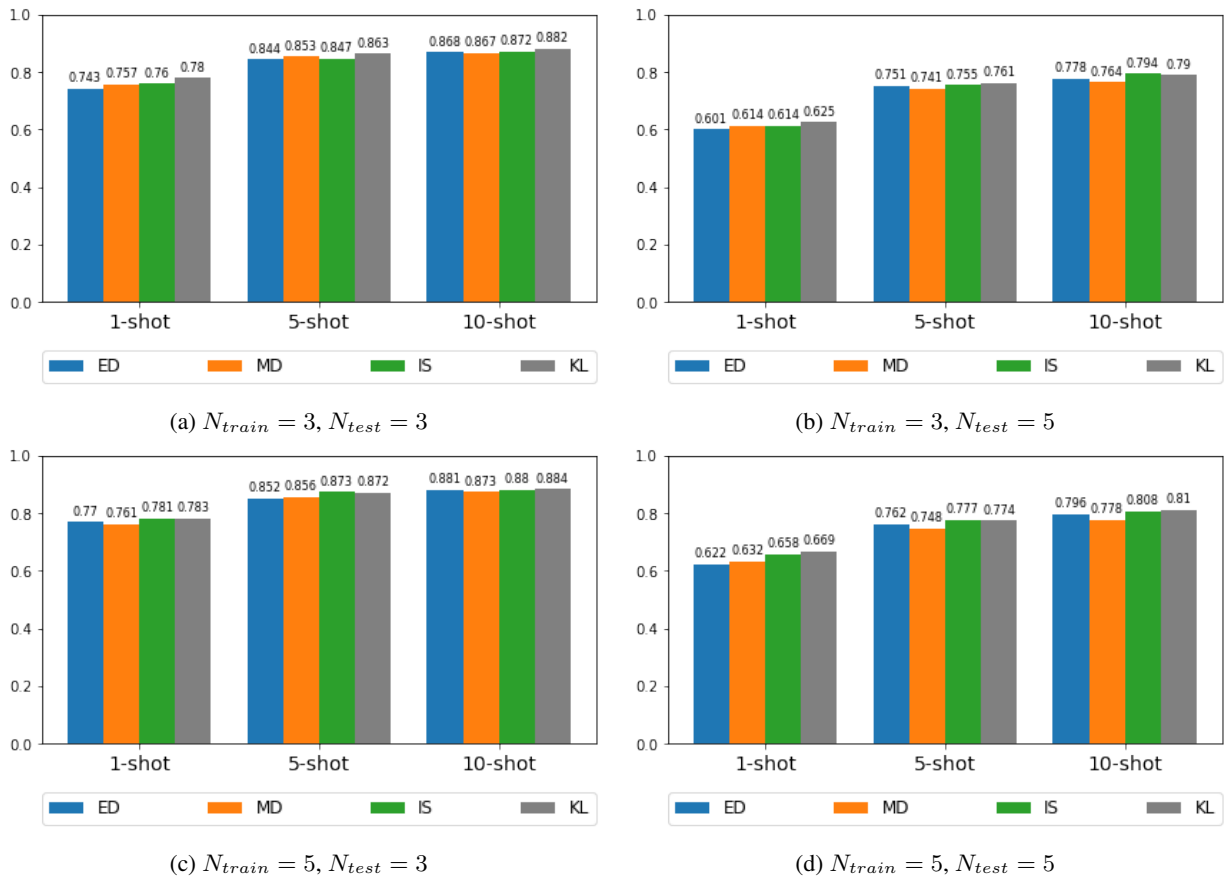


Figura 3.7: Comparação das divergências de Bregman para a fase adulta. ED: Distância Euclidiana. MD: Distância de Mahalanobis. KL: Divergência de Kullback-Leibler. IS: Divergência de Itakura-Saito.

mais robusta em alguns casos. Na seção a seguir, apontamos um possível fator que explica em que circunstância a robustez da distância de Mahalanobis pode ser melhorada.

### 3.5.3.2 Distância de Mahalanobis

Conforme a metodologia adotada nesse trabalho, a distância de Mahalanobis representa uma medida de distância entre os pontos de consulta e cada protótipo dada a distribuição dos protótipos de classes dentro de uma tarefa. Dessa forma, a distância de Mahalanobis depende da matriz de variância-covariância entre os protótipos, comum para todas as medições de distância em uma tarefa.

A distância de Mahalanobis é uma medida robusta quando temos variáveis que possuem alguma correlação linear. Entretanto, os dados de características de imagens são altamente aleatórios e, portanto, pode não haver correlação definida entre quaisquer *embeddings*, independentemente de pertencerem a mesma classe ou não. Por esse motivo, considerar a variabilidade conjunta entre os protótipos, pode não ser a melhor opção para os dados IP-FSL.

O número de exemplos dentro do conjunto de protótipos de uma tarefa é naturalmente muito menor do que a dimensão dos vetores no espaço de *embeddings*. Por esse motivo, a matriz de co-

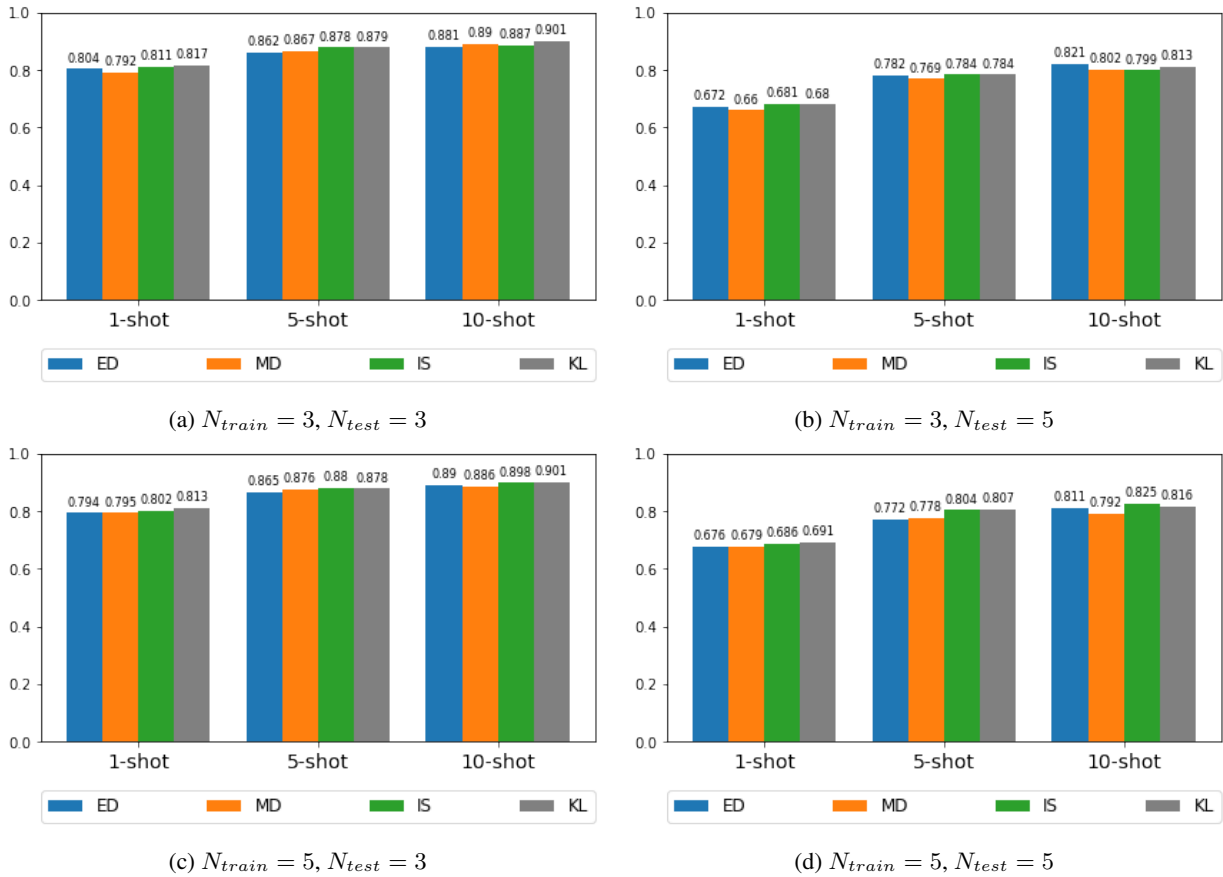


Figura 3.8: Comparação das divergências de Bregman para a fase primária. ED: Distância Euclidiana. MD: Distância de Mahalanobis. KL: Divergência de Kullback-Leibler. IS: Divergência de Itakura-Saito.

variância usada na Equação 3.3 pode ser melhor estimada por meio da combinação de duas outras matrizes de covariância, calculadas a partir dos *embeddings* de classe e os *embeddings* de tarefa respectivamente (BATENI et al., 2020). No entanto, o cálculo da covariância exige um custo de tempo considerável e, portanto, o tempo total de treinamento do modelo pode aumentar expressivamente com o cálculo de outras duas matrizes, dado que os modelos de FSL são treinados a partir de uma grande quantidade de tarefas.

Os experimentos nesse trabalho focaram na redução de tempo de processamento para cálculo da distância de Mahalanobis, por isso, a covariância foi calculada a partir dos protótipos de classe. Embora a distância de Mahalanobis não tenha alcançado resultados gerais superiores aos das outras divergências, o método aqui proposto produziu resultados competitivos. Os resultados mostraram que a distância de Mahalanobis produziu resultados ligeiramente inferiores aos das divergências KL e IS. Comparando-se à distância Euclidiana, os resultados oscilaram e não obedeceram a uma regra de configuração. Portanto, a melhor métrica pode ser selecionada com base na eficiência do processo de treinamento, considerando, por exemplo, o tempo gasto e a capacidade de processamento disponível.



### 3.5.3.3 Divergência KL

Conforme os esclarecimentos sobre as divergências de Bregman na Seção 3.1, a divergência KL não é considerada uma métrica de distância, tendo em vista que é assimétrica e não cumpre a desigualdade triangular. Por outro lado, a divergência de Jensen-Shannon (ENDRES; SCHINDELIN, 2003) estende a divergência KL para calcular uma pontuação simétrica, podendo ser considerada uma métrica de distância.

A divergência KL calcula uma pontuação que quantifica o quanto uma distribuição de probabilidade se diferencia de outra. Essa abordagem foi utilizada nesse trabalho para diferenciar as distribuições de probabilidade geradas a partir dos *embeddings* de consulta e dos protótipos. Dadas essas distribuições, a entropia relativa do protótipo  $\mathbf{y}$  para um ponto de consulta  $\mathbf{x}$  foi calculada como  $KL(\mathbf{x}||\mathbf{y})$ . Um valor de divergência baixo, ou seja, quando  $KL(\mathbf{x}||\mathbf{y})$  se aproxima de zero, indica que as distribuições são semelhantes. Por outro lado, conforme os valores se distanciam positivamente de zero maior será a diferença entre as distribuições.

Os resultados mostraram que a divergência KL é uma medida robusta para classificação dos dados IP-FSL. Na maioria das configurações experimentais, a KL produziu resultados superiores aos das demais divergências investigadas. Nos dados de insetos adultos, por exemplo, ela foi superada apenas pela divergência IS, em 3 dos 12 experimentos realizados nessa categoria, com diferença de no máximo 0,4 ponto percentual, foram eles:

- Figura 3.7b:  $N_{train} = 3$ ,  $N_{test} = 5$  e *10-shot*;
- Figura 3.7c:  $N_{train} = 5$ ,  $N_{test} = 3$  e *5-shot*;
- Figura 3.7d:  $N_{train} = 5$ ,  $N_{test} = 5$  e *5-shot*.

Nos dados de insetos da fase primária, KL também foi superada pela mesma divergência, IS, em outros 3 casos, com diferença máxima de 0,9 ponto percentual:

- Figura 3.8b:  $N_{train} = 3$ ,  $N_{test} = 5$  e  $K = 1$ ;
- Figura 3.8c:  $N_{train} = 5$ ,  $N_{test} = 3$  e  $K = 5$ ;
- Figura 3.8d:  $N_{train} = 5$ ,  $N_{test} = 5$  e  $K = 10$ .

Além desses casos citados, a KL foi superada pela distância Euclidiana em um ponto isolado na classificação da fase primária (Figura 3.8b:  $N_{train} = 3$ ,  $N_{test} = 5$  e  $K = 10$ ). Apesar desses pontos específicos, KL apresentou melhores resultados gerais em relação as demais divergências.

### 3.5.3.4 Divergência IS

Em geral, a divergência IS produziu resultados similares aos da divergência KL, na maioria dos experimentos. Embora algumas flutuações dos valores tenham sido constatadas em relação

às outras divergências, considera-se também uma medida robusta para quantificar a diferença entre os vetores de probabilidade dos dados de insetos. Entretanto, os resultados de todas as divergências são muito aproximados e provavelmente uma análise aprofundada seja necessária para esclarecimentos sobre o rendimento de cada divergência investigada nesse trabalho. Por esse motivo, analisamos os resultados sob aspectos específicos que serão detalhados nas próximas seções.

### 3.5.3.5 Análise de sensibilidade

Tendo em vista a proximidade entre os valores dos resultados, realizou-se uma análise de sensibilidade das divergências em lidar com uma tarefa de classificação considerada de maior complexidade. Essa análise funciona como uma espécie de lupa para a visualização dos resultados em dados específicos. A análise focou em um ponto crítico dos dados de insetos, no qual o elevado grau de semelhança visual entre algumas classes torna o processo de classificação ainda mais desafiador.

Cinco classes de insetos adultos, com características visuais relativamente semelhantes, foram selecionadas para compor o conjunto de teste (*target set*). Tais classes pertencem às espécies de gorgulho, e são comumente encontradas em culturas de arroz, beterraba, alfafa e manga. Neste teste, utilizamos as configurações *N-way-K-shot* mais promissoras. O modelo foi treinado utilizando as demais classes dos dados de insetos adultos.

Os resultados apresentados na Figura 3.9, mostram haver uma degradação da acurácia em comparação com a classificação de insetos adultos mostrada nas Figuras 3.7a e 3.7c, usando os mesmos parâmetros de aprendizagem e de teste. Isso mostra que a dificuldade de aprendizagem é maior quando classes muito semelhantes são exploradas dentro de uma tarefa de classificação, tornando o processo mais desafiador como relatado anteriormente. Apesar disso, observou-se que mesmo em uma situação de grande semelhança entre classes de uma tarefa, as divergências de Bregman ainda apresentam desempenho relevante para a classificação de insetos.

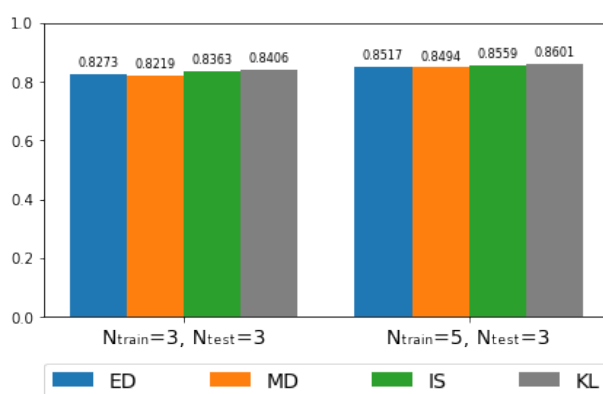


Figura 3.9: Acurácia para classes similares de insetos adultos usando *10-shot*.

Observou-se que à medida que a diversidade do conjunto de dados aumenta, isto é, quando

são exploradas classes semelhantes e não semelhantes nas tarefas de classificação, o modelo se torna mais propenso a classificar corretamente. Embora os desempenhos das divergências sejam relativamente semelhantes, os resultados mostrados na Figura 3.9 indicam que a divergência KL mostrou maior efetividade. Além disso, esses resultados são parâmetros importantes para a tomada de decisão sobre a escolha da divergência em FSL.

### 3.5.3.6 Análise empírica

A análise empírica foi realizada com intuito de demonstrar o desempenho de classificação em duas perspectivas: 1) o desempenho médio dos modelos de FSL nos pontos com maiores índices de classificação correta dos dados e, 2) demonstrar a eficiência dos modelos com respeito ao tempo de treinamento e a acurácia alcançada.

Para análise de desempenho médio, foram considerados os parâmetros  $N$ -way de treinamento e teste com maiores possibilidades de acerto ( $N_{train} = 3, N_{test} = 3$  e  $N_{train} = 5, N_{test} = 3$ ). Dessa forma, obteve-se o desempenho médio das divergências calculando-se a média das acurácias, em relação a  $K$ -shot, dos resultados contidos nas Figuras 3.7 e 3.8.

A Figura 3.10 mostra os resultados médios usando as configurações  $N$ -way que apresentaram os melhores desempenhos de classificação de insetos adultos (Figuras 3.7a e 3.7c) e da fase primária (Figuras 3.8a e 3.8c). Embora as divergências simétricas (ED e MD) sejam usadas com frequência em FSL, foi demonstrado aqui que as divergências assimétricas (KL e IS), mesmo não sendo consideradas métricas de distância, são soluções competitivas para medição de similaridade em modelos de FSL, demonstrando melhores resultados em relação às simétricas, conforme mostraram as Figuras 3.9 e 3.10.

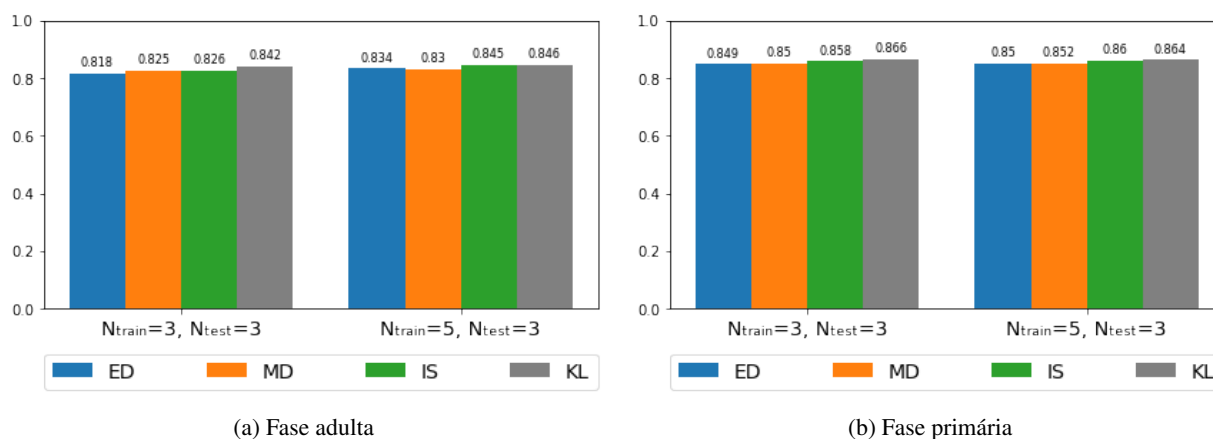


Figura 3.10: Média das acurácias usando as configurações mais promissoras para classificação *few-shot*.

Além do conhecimento sobre os parâmetros com maior possibilidade de acerto em FSL, o tempo de treinamento relacionado à cada uma das divergências pode ser outro fator de decisão. Embora possa haver pouca diferença no tempo de processamento entre algumas divergências, essa diferença pode ser expressiva após milhares de tarefas de classificação. Recuperamos o tempo de treinamento referente à cada divergência durante o treinamento de uma época usando

as configurações *3-way*, *10-shot* e  $q = 5$ . Utilizamos o experimento com o conjunto de dados de insetos adultos, entretanto, a escolha dos dados é redundante para análise de tempo, visto que as imagens de entrada possuem as mesmas dimensões ( $96 \times 96 \times 3$ ) para extração de recursos usando o mesmo modelo CNN.

Os tempos de treinamento apresentados na Figura 3.11 mostram que a distância Euclidiana e as divergências KL e IS produzem tempos de processamento semelhantes. Por outro lado, a distância de Mahalanobis requer um tempo elevado. Presume-se, portanto, que o tempo gasto é proporcional à complexidade da equação correspondente. Com isso, a razão do elevado tempo de treinamento associado à distância de Mahalanobis, é explicada pelo custo computacional necessário para o cálculo da matriz de covariância, tornando a razão acurácia-tempo muito baixa para essa distância. Mesmo com um tempo de treinamento elevado, a distância de Mahalanobis não produziu um desempenho consideravelmente alto em relação às outras divergências.

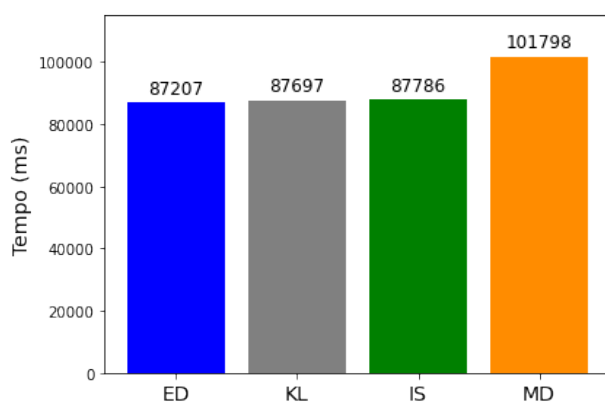


Figura 3.11: Tempo gasto após um episódio de treinamento usando *3-way*, *10-shot* e  $q = 5$ .

Em um processo de seleção da divergência a ser usada no modelo de FSL, se o tempo for um fator crucial em uma determinada aplicação, de modo que se sobreponha a melhor acurácia, os métodos com equações mais simples são os mais aconselháveis. Por outro lado, as divergências com maior relação acurácia-tempo podem ser vistas como abordagens relevantes na busca de um equilíbrio entre custo de processamento e acurácia desejada. A Figura 3.12, construída a partir dos dados das Figuras 3.10 e 3.11, mostra os resultados globais dessa relação para cada divergência de Bregman investigada. Os gráficos da Figura 3.12 mostraram que a distância de Mahalanobis apresentou o menor índice de acertos em relação ao tempo de treinamento. As demais divergências produziram taxas relativamente semelhantes, com a divergência KL apresentando a melhor relação.

Outros fatores também contribuem para o aumento de desempenho de classificação do modelo em termos gerais, como o número de episódios que compõem uma época, além da quantidade de épocas de treinamento. A medida que o número de tarefas (episódios) de classificação aumenta, mais informações são aprendidas até o limite de aprendizado fornecido pela capacidade do modelo e pela quantidade de dados. Por esse motivo, é importante que a quantidade de tarefas na fase de treinamento seja suficiente para o aprendizado quando um conjunto de dados possui mui-

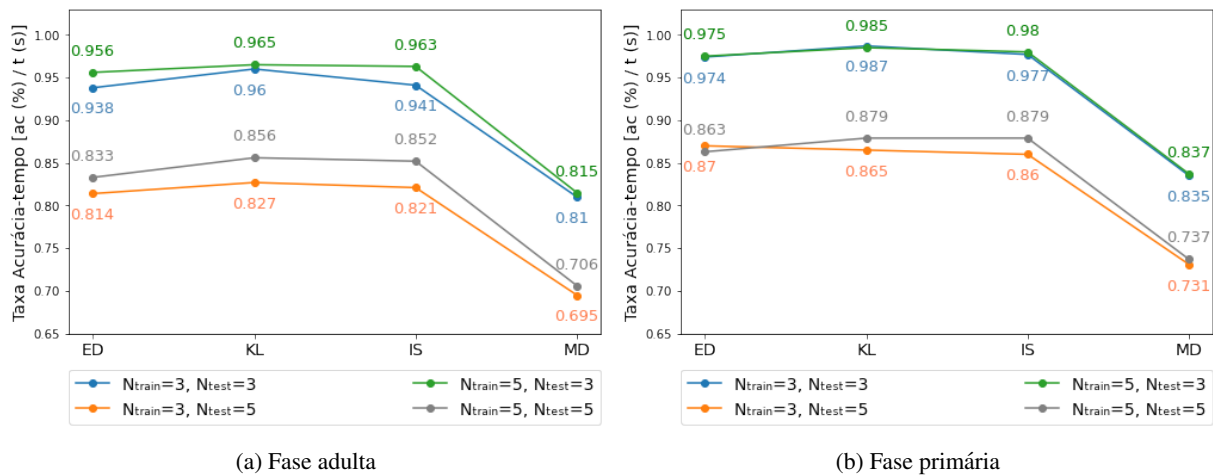


Figura 3.12: Taxa Acurácia-tempo.

tas classes. Dessa forma, o modelo explora mais classes durante o aprendizado, mitigando as chances de erros de classificação durante os testes.

### 3.6 CONCLUSÕES DO CAPÍTULO

Este capítulo forneceu o primeiro estudo sobre a comparação das divergências de Bregman para classificação de insetos usando aprendizagem *few-shot*. Um conjunto de dados de insetos da agricultura (IP-FSL) contendo 142 classes, com categorias separadas por estágios de maturidade, fase adulta (97) e fase primária (45), foi proposto para o treinamento e teste dos modelos.

Realizou-se aqui uma cadeia de 48 experimentos para análise dos parâmetros  $N$ -way- $K$ -shot e das divergências em cada categoria dos dados IP-FSL, totalizando 96 experimentos. Os resultados mostraram que as configurações definidas como  $N_{train} > N_{test}$  e  $N_{train} = N_{test} = 3$  produziram os melhores resultados. Isso mostrou que a facilidade de aprendizagem e de classificação correta nos testes depende do número de classes exploradas nas tarefas. Quanto maior o número de classes, maior a possibilidade de ocorrerem erros de classificação.

As divergências mostraram resultados semelhantes entre si, e a escolha do método depende de alguns fatores como, por exemplo, a relação acurácia-tempo de cada divergência. Os melhores resultados foram associados às divergências com equações mais simples nessa análise. Em contraste, os classificadores baseados nas divergências assimétricas produziram medidas de similaridade mais relevantes, com os melhores resultados usando a divergência de Kullback-Leibler, mostrando que essas abordagens, assim como as divergências simétricas, são soluções competitivas para a classificação de insetos.

O reconhecimento visual usando poucos dados é visto como um desafio em aprendizagem de máquinas. Realizar o melhor uso da CNN para extração de recursos relevantes de imagens e definir como usá-los para classificar objetos não é uma questão trivial. O próximo capítulo

apresenta um modelo aprimorado de aprendizagem *few-shot*, construído com base na questão não trivial citada que será discutida em uma aplicação específica de classificação de insetos em agricultura.

## 4 APRENDIZAGEM *FEW-SHOT* VIA FUSÃO DE RECURSOS E ENTROPIA RELATIVA

Este capítulo apresenta um estudo sobre melhorias de desempenho durante a extração de características de imagens e de classificação a partir de uma medida efetiva de similaridade. A abordagem aqui apresentada inclui um mecanismo que estende a arquitetura CNN usada no capítulo anterior para extração e fusão de características de imagens, aumentando a representatividade dos vetores de características. Um conjunto de dados de insetos das culturas de milho, com categorias de pragas e insetos benéficos (incluindo predadores e parasitoides), foi construído para teste do modelo proposto. Experimentos foram realizados para comparação com as redes ResNet50, VGG16 e MobileNetv2, usadas como *backbones* para extração de recursos de imagens. Os resultados mostraram o desempenho superior do modelo apresentado.

### 4.1 O MODELO PROPOSTO

Um dos grandes desafios encontrados nos modelos de aprendizagem de máquinas está ligado a disponibilidade de dados suficientes para o aprendizado. Na aprendizagem profunda, os bancos de dados são normalmente constituídos de milhares de imagens, fornecendo quantidade expressiva de informações sobre classes. Já na aprendizagem *few-shot*, existem poucas informações disponíveis para a construção de conhecimento sobre classes de dados. Por esse motivo, quanto maior a quantidade de características relevantes de imagem extraídas, mais consistente pode ser o conhecimento adquirido pelo modelo. Entretanto, definir quais informações de imagens extrair e como usá-las é uma tarefa desafiadora.

Trabalhos anteriores basearam-se na extração de recursos de textura e de cores de imagens segmentadas usando os métodos de Matriz de Coocorrência de Nível de Cinza (GLCM) e Histograma de Cores para tarefas de classificação em agricultura (ES-SAADY et al., 2016; ISLAM et al., 2017). No entanto, algoritmos de aprendizagem profunda baseados em CNN, demonstraram melhores resultados tendo em vista o processo de extração automática, sem a necessidade de pré-processamento de imagens, e a acurácia alcançada. Esses algoritmos mostraram bom desempenho em tarefas agrícolas como classificação de insetos (THENMOZHI; REDDY, 2019) e doenças de plantas (ZHANG et al., 2018).

Na maioria dos trabalhos envolvendo aprendizagem *few-shot*, a extração de características globais de imagens é uma tarefa comumente usada para a construção de vetores representativos de imagens. Entretanto, quando as características locais não são consideradas, pode haver perda de informações importantes. No cenário de poucos dados, as características locais podem fornecer informações relevantes sobre imagens e, negligenciá-las durante a extração de características,

pode enfraquecer a representatividade dos vetores gerados e penalizar a medição de similaridade.

Diante do contexto apresentado, o método proposto nesse capítulo se baseia na hipótese de que a fusão de características locais e globais, vindas das camadas internas da CNN, pode ser vista como um método mais eficiente para a construção de vetores representativos em relação ao uso de apenas das informações globais. Com a fusão desses recursos visuais internos, espera-se mitigar a perda de informações para a construção de vetores que representem fortemente as classes de dados. Além disso, consideramos outra hipótese de que o desempenho do classificador pode ser melhorado neste cenário usando uma medida de divergência, como a entropia relativa, em vez de uma métrica de distância como é normalmente adotada na literatura.

A extração de características internas da CNN é um método promissor usado em outros trabalhos de aprendizagem profunda, tais como (FU et al., 2019; AMIN et al., 2019). *Feature Pyramid Networks* (FPN) (LIN et al., 2017), por exemplo, é um dos métodos mais conhecidos na literatura. FPN funde recursos multicamadas para predição de caixas delimitadoras (*bounding boxes*) para detecção de objetos. Já Kim e Chi (2021) usaram um mecanismo de atenção para selecionar recursos multicamadas e fundi-los para classificação de imagens de sensoriamento remoto usando *few-shot*. Outros autores, como (DING et al., 2019; JIANG et al., 2020), apostaram na fusão de recursos de várias camadas da CNN para a geração de características de relação, usadas nas redes de Relação (*Relation networks*). A nossa abordagem integra um método específico de extração de recursos multicamadas ao modelo *few-shot*, baseada em uma rede CNN rasa, e propõe uma metodologia simplificada de fusão desses recursos para a formação de representantes de classes ricos em informações.

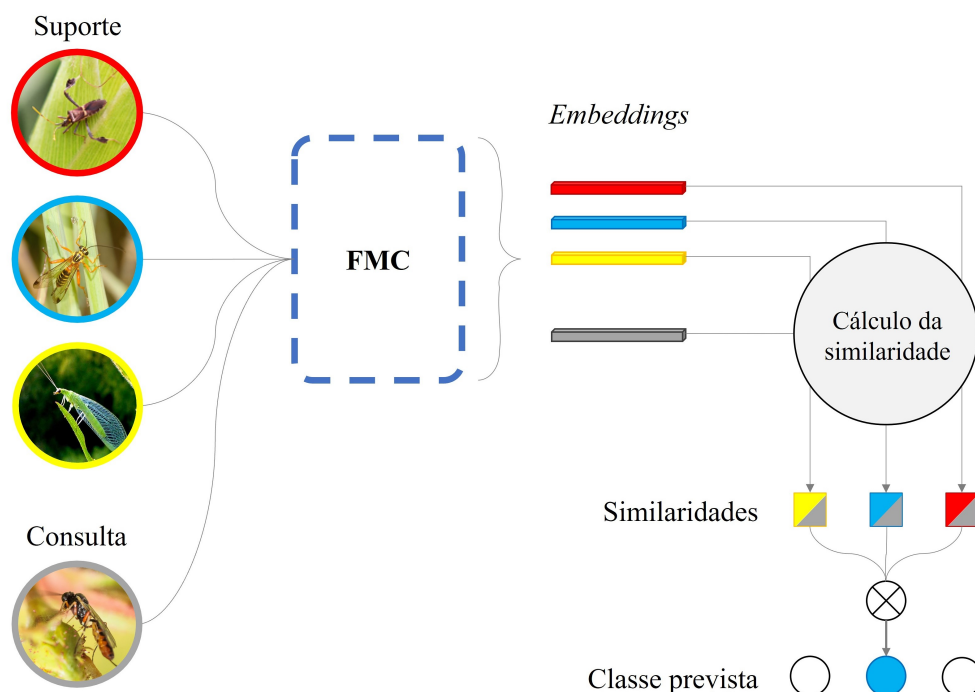


Figura 4.1: Arquitetura do modelo proposto.

A Figura 4.1 mostra a arquitetura do modelo proposto, enfatizando os dois principais módulos



de sua composição: um bloco de extração e Fusão de recursos Multicamada (FMC), e um classificador baseado na divergência de Kullback-Leibler, também conhecida como Entropia Relativa, ou a distância Euclidiana.

A estrutura FMC foi construída para a geração de vetores ricos em informações, mesmo sendo uma estrutura baseada em uma rede considerada simples (isto é, rasa). O nosso modelo funde informações locais, medianas da rede e globais de todas as imagens de uma tarefa de classificação, possibilitando a formação de *embeddings* e protótipos ricos em informações de classes. Na próxima seção, são apresentados os detalhes da FMC.

### 4.1.1 Extração de características de imagem

Trabalhos recentes mostraram que a CNN é um método consolidado para extração de recursos discriminativos de imagens, mesmo usando uma arquitetura de rede rasa (FU et al., 2019; LI; NIE; CHAO, 2020). A compreensão do funcionamento por trás das camadas convolucionais de uma CNN foi apresentada intuitivamente por Zeiler e Fergus (2014). Conforme a Figura 4.2, as primeiras camadas caracterizam um detector de recursos de baixo nível, como bordas, por exemplo. Em vez disso, as camadas mais profundas da rede detectam recursos de alto nível, como rostos e flores, por exemplo, nesse caso específico.

A proposta nesse capítulo consiste em realizar a fusão desses recursos do baixo ao alto nível para a geração de vetores consistentes de informações. Conforme mostrado na Figura 4.3, a arquitetura proposta consiste em quatro blocos convolucionais. Cada bloco contém uma camada convolucional com 64 filtros  $3 \times 3$ , normalização, Unidade Linear Retificada (ReLU) e *Max Pooling*  $3 \times 3$ . A Figura 4.3 mostra os detalhes do nosso método. Após a saída de cada bloco, são adicionadas camadas de *Max Pooling* e *Flatten* para o fornecimento de vetores de 64 dimensões que representam cada bloco. A Figura B.1 (Apêndice B), mostra um resumo de todas as camadas do modelo de acordo com o *backbone* principal e suas ramificações  $f_x$  ( $x = 1 : 4$ ).

A saída da estrutura de fusão de recursos (SFR) resulta da média dos vetores representantes, conforme mostra a Figura 4.4. Em termos formais, a partir de uma respectiva função de extração de recursos  $f_\phi$  com parâmetros de aprendizagem  $\phi$ , a fusão de recursos calcula a média dos vetores representativos  $f$  das camadas de  $f_\phi$ , conforme a Equação 4.1.

$$\text{SFR}^n = \frac{1}{4} \sum_{i=1}^4 f_i(x) \quad (4.1)$$

Em que  $\text{SFR}^n$  representa o *embedding* da imagem  $x$  na classe  $n$ .

O objetivo da estrutura FMC em conjunto com a SFR, é aprender representantes consistentes que agreguem o máximo de informações possíveis sobre a imagem. Espera-se, portanto, que o mínimo de informação seja perdido durante o processo. Todas as imagens na tarefa de classificação, isto é, as imagens de suporte e consulta, passam pela estrutura de extração e fusão

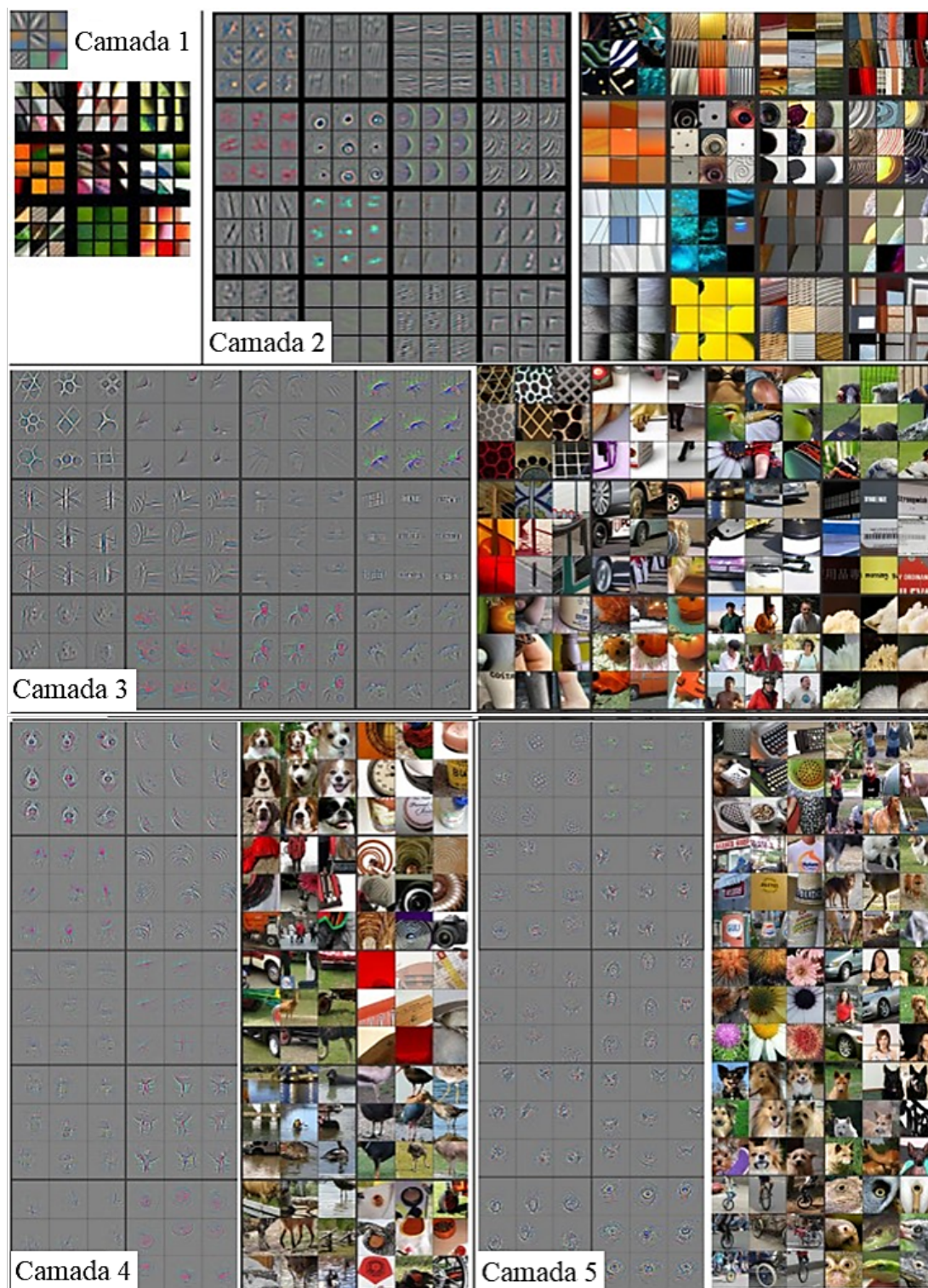
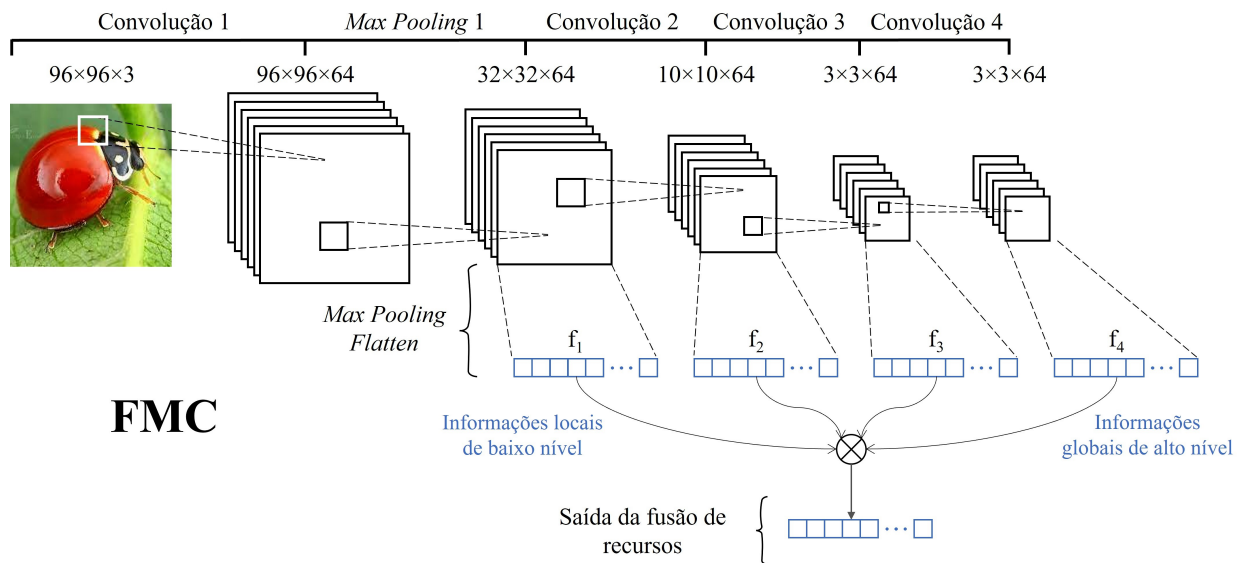


Figura 4.2: Visualização e entendimento das camadas da CNN.  
 Fonte: Adaptado de (ZEILER; FERGUS, 2014).



## FMC

Figura 4.3: O modelo proposto para extração e fusão de recursos multicamadas da CNN (FMC).

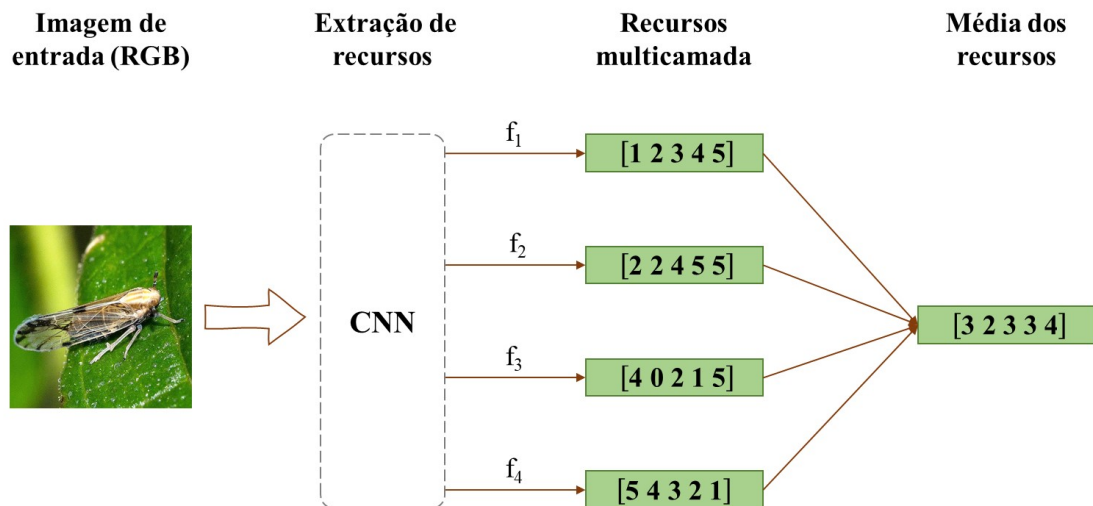


Figura 4.4: Exemplo representativo de extração e fusão de recursos multicamada (FMC) de uma imagem.

de recursos.

Para a obtenção dos vetores representativos de classe, a geração de protótipos é realizada sobre os *embeddings* de suporte após a SFR. Cada protótipo corresponde ao centroide do *cluster* de *embeddings* de classe, calculado conforme a Equação 4.2.

$$c_n = \frac{1}{K_n} \sum_{i=1}^{K_n} \text{FFO}^n(\mathbf{x}_i) \quad (4.2)$$

Em que  $c_n$  representa o centroide da classe  $n$ .

A classificação das imagens de consulta em uma tarefa é realizada a partir de uma medida de dissimilaridade entre os protótipos e os *embeddings* de consulta. A abordagem adotada será detalhada na próxima seção.

## 4.1.2 Medição de similaridade

A escolha de uma medida de similaridade robusta em *few-shot* é tão importante quanto a adoção de um método eficiente para extração de recursos de imagem. O termo 'robustez' expressado aqui se refere a capacidade de uma medida de similaridade em lidar com os vetores construídos a partir de informações de baixo e alto níveis.

As métricas de distância foram amplamente adotadas na literatura para comparação de *embeddings* em *few-shot*. Métricas como a diferença absoluta (YANG et al., 2020), a distância Euclidiana (SNELL; SWERSKY; ZEMEL, 2017), a distância de Mahalanobis (NUTHALAPATI; TUNGA, 2021) e a similaridade do cosseno (VINYALS et al., 2016) são abordagens comuns na maioria dos trabalhos. No entanto, outras medidas de similaridade, incluindo as divergências, não receberam atenção em aplicações de *few-shot*, possivelmente por se tratar de uma abordagem de aprendizagem relativamente recente.

Para a escolha do método de similaridade do modelo proposto nesse capítulo, consideramos os resultados alcançados no Capítulo 3. Tais resultados mostraram que as divergências assimétricas, embora ainda não tenham sido amplamente adotadas na literatura em *few-shot*, são métodos competitivos para medição de similaridade nessa aplicação. Portanto, além da distância Euclidiana, já consolidada em muitos trabalhos, adotamos a Entropia relativa como função de similaridade neste capítulo para obtenção da melhor diferenciação entre *embeddings*. Espera-se que essas divergências sejam escolhas promissoras para lidar com a comparação de *embeddings* gerados pelo nosso modelo de fusão de recursos de imagens.

Como explicado no Capítulo 3, a distância Euclidiana calcula a distância quadrática, simétrica, entre os pontos dos vetores comparados. Já a Entropia relativa é uma medida assimétrica que quantifica a diferença entre duas distribuições de probabilidade, calculada pela Equação 3.4. O logaritmo presente na equação pode ser de base 2 ou base  $e$  para medir a divergência em unidades de bits ou nats respectivamente. Se o resultado alcançado no cálculo da divergência entre duas

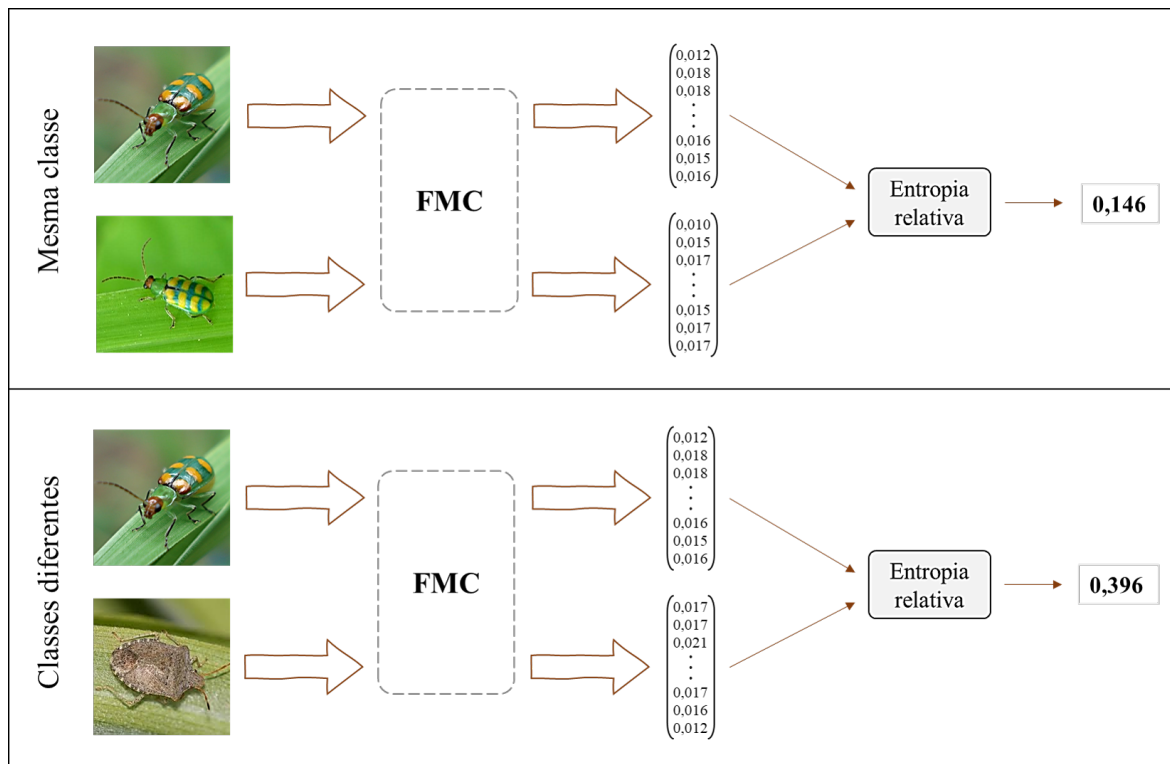


Figura 4.5: Desempenho esperado no cálculo da divergência. Resultados obtidos usando o modelo treinado.

distribuições for um valor próximo de zero, significa que elas são muito semelhantes e possuem alta probabilidade de pertencerem à mesma classe. Por outro lado, se o valor alcançado se distancia de zero positivamente, isso indica uma alta divergência entre elas e a probabilidade é baixa. O desempenho esperado na obtenção da Entropia relativa entre *embeddings* pode ser explicado conforme mostra a Figura 4.5, em que usamos a rede CNN treinada para extração de características dos exemplos. Para o cálculo da entropia entre os protótipos e *embeddings*, adotamos a mesma abordagem sugerida no Capítulo 3.

## 4.2 CONJUNTO DE DADOS DE INSETOS DO MILHO

São necessárias apenas algumas amostras de dados para o treinamento do modelo proposto nesse capítulo. Construímos um pequeno conjunto de dados de insetos para os testes do modelo. Os dados são específicos das culturas de milho, abrangendo insetos benéficos (predadores e parasitoides) e pragas. Dois motivos foram fundamentais para a escolha de dados de insetos da agricultura para os experimentos desse capítulo: 1) a contribuição para o reconhecimento automático de insetos na agricultura, e 2) devido ao reconhecimento de insetos ser uma tarefa não trivial, inclusive para seres humanos, dada a semelhança entre as espécies.

O “Panorama Fitossanitário da cultura do milho”, Embrapa Milho e Sorgo, disponível no endereço eletrônico: <http://panorama.cnpm.embrapa.br/insetos-praga>, e também o “Guia ilus-



Tabela 4.1: Classes do conjunto de dados de insetos das culturas de milho.

Conjunto de origem			Conjunto de destino		
Rótulo	Nome do inseto	Tipo	Rótulo	Nome do inseto	Tipo
1	<i>Dalbulus maidis</i>	Praga	15	<i>Astylus variegatus</i>	Praga
2	<i>Deois flavopicta</i>	Praga	16	<i>Delia spp</i>	Praga
3	<i>Diatraea saccharalis</i>	Praga	17	<i>Diabrotica speciosa</i>	Praga
4	<i>Elasmopalpus lignosellus</i>	Praga	18	<i>Diloboderus abderus</i>	Praga
5	<i>Helicoverpa zea</i>	Praga	19	<i>Euxesta spp</i>	Praga
6	<i>Leptoglossus zonatus</i>	Praga	20	<i>Frankliniella williamsi</i>	Praga
7	<i>Mocis latipes</i>	Praga	21	<i>Rhopalosiphum maidis</i>	Praga
8	<i>Peregrinus maidis</i>	Praga	22	<i>Scaptocoris castanea</i>	Praga
9	<i>Spodoptera frugiperda</i>	Praga	23	<i>Campoletis flavicincta</i>	Parasitoide
10	<i>Cycloneda sanguinea</i>	Predador	24	<i>Ceraeochrysa sp.</i>	Predador
11	<i>Geocoris sp.</i>	Predador	25	<i>Doru luteipes</i>	Predador
12	<i>Hippodamia convergens</i>	Predador	26	<i>Lysiphelus testaceipes</i>	Parasitoide
13	<i>Orius sp.</i>	Predador	27	<i>Telenomus remus</i>	Parasitoide
14	<i>Podisus sp.</i>	Predador	28	<i>Trichogramma pretiosum</i>	Parasitoide

trado de pragas e insetos benéficos do milho e sorgo” (PINTO; PARRA; OLIVEIRA, 2004), foram referências importantes para a construção do conjunto de dados. O conjunto é composto de 28 classes de insetos, sendo 17 consideradas pragas do milho, e 11 classes de alguns dos principais inimigos naturais associados às pragas dessa cultura. Cada classe de inseto é composta de 15 imagens, suficiente para treinamento e teste do modelo de *few-shot*. Dessa forma, todo o conjunto de dados conta com 420 imagens.

Os dados foram divididos em conjuntos chamados aqui de origem (*Source set*) e de destino (*Target set*) para treinamento e teste do modelo baseado em meta-aprendizagem, conforme mostra a Tabela 4.1. Nessa tabela, todas as informações do conjunto de dados são apresentadas. A teoria sobre os conjuntos de origem e destino foi explicada com detalhes no Capítulo 3. Uma amostra de cada classe de inseto pode ser visualizada na Figura 4.6, onde as imagens das classes foram agrupadas por categoria (pragas e insetos benéficos).

### 4.3 DETALHES EXPERIMENTAIS

Nesse capítulo uma cadeia experimental foi organizada para a análise do desempenho do modelo aqui proposto na classificação de insetos da cultura do milho. Os experimentos consistiram em 3 cenários, 1) treinamento e teste do modelo extraindo apenas características globais de imagens (corresponde ao recurso  $f_4$  da Figura 4.3, chamado aqui *Backbone<sup>f4</sup>*), 2) treinamento e teste do modelo usando o FMC proposto, e 3) avaliamos três redes CNN da literatura como *backbones* extratores de recursos de imagens para a comparação dos resultados, cujas camadas de classificação (*Fully Connected layers - FC*) foram removidas, são elas: ResNet50 (HE et al., 2016), VGG16 (SIMONYAN; ZISSERMAN, 2015) e MobileNetv2 (SANDLER et al., 2018). A Tabela 4.2 mostra a quantidade de parâmetros em cada *backbone*.

Da mesma forma como foi realizado no Capítulo 3 para a preparação do conjunto de dados,



Figura 4.6: Amostras de imagens de cada classe dos dados.

foram realizadas operações morfológicas nas imagens, como redimensionamento e rotação, para padronização e aumento do número de classes. As imagens foram redimensionadas para a forma  $96 \times 96 \times 3$  e novas classes foram criadas aplicando-se rotações de  $90^\circ$ ,  $180^\circ$  e  $270^\circ$ .

A configuração dos parâmetros *N-way-K-shot* pode ser realizada de diferentes maneiras, conforme visto no Capítulo 3, e também em outros trabalhos (JAMAL; QI, 2019; LI; YANG, 2021; LI; CHAO, 2021; NUTHALAPATI; TUNGA, 2021). No entanto, todos os modelos investigados neste capítulo foram treinados sob a mesma configuração, incluindo o parâmetro das imagens de consulta  $q$ , definidos como: *5-way*, *1-shot* e *5-shot*, e  $q = 5$ . As imagens de consulta dentro de uma tarefa são classificadas de acordo com uma distribuição de probabilidade sobre as di-

Tabela 4.2: Número de parâmetros em cada modelo.

<i>Backbone</i>	Parâmetros ( $10^3$ )
VGG16	138.365,99
Resnet50	25.557,03
Mobilenetv2	3.504,87
Recurso $f_1$ da FMC	1,92
Recurso $f_2$ da FMC	38,98
Recurso $f_3$ da FMC	76,03
Recurso $f_4$ da FMC	113,09
Total da FMC	230,02

vergências obtida pelo *softmax* conforme a Equação 3.8. Além disso, o nosso modelo também aprende minimizando a probabilidade de logaritmo negativo  $J(\phi) = -\log_{\phi}(y = n|\mathbf{x})$ .

Os modelos foram treinados usando os dados de origem por 100 épocas. Cada época foi definida para executar 150 tarefas de classificação (ou 150 episódios). A taxa de aprendizagem inicial foi definida como  $10^{-3}$  sendo dividida por dois após 20 épocas de aprendizagem. Os testes foram realizados usando o conjunto de destino e a acurácia foi obtida pela média de 2000 episódios de teste. A acurácia indica um desempenho geral do modelo, ou seja, dentre todas as classificações realizadas, ela fornece uma medida de quantas amostras o modelo classificou corretamente, conforme a Equação 4.3.

$$acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (4.3)$$

Em que:

$acc$  — é a própria acurácia.

$TP$  — indica a quantidade de amostras positivas classificadas corretamente.

$TN$  — a quantidade de amostras negativas classificadas corretamente.

$FP$  — a quantidade de amostras negativas classificadas de maneira incorreta como positivas.

$FN$  — a quantidade de amostras positivas classificadas de maneira incorreta como negativas.

## 4.4 RESULTADOS E DISCUSSÕES

A construção e preparação de conjuntos de dados é uma tarefa que normalmente demanda um tempo expressivo em aprendizagem de máquinas. Na maioria dos métodos de aprendizagem, a disponibilidade de grandes conjuntos de dados é necessária para aprendizado dos modelos, afinal, aprender a partir de poucos dados é uma tarefa desafiadora em reconhecimento visual. A aprendizagem *few-shot*, entretanto, não necessita de grandes conjuntos de dados de treinamento, facilitando a aquisição de dados para aplicações específicas e relevantes.

Este capítulo forneceu um conjunto de dados desafiador de insetos da cultura de milho para aprendizagem *few-shot*, com separação exclusiva de categorias de insetos benéficos (incluindo predadores e parasitoides) e pragas. O conjunto abrange 28 classes de insetos contendo 15 imagens em cada uma delas, dividido consoante a Tabela 4.1 para avaliação do modelo proposto nesse capítulo. Conforme mostra a tabela de dados, não há relação entre o conjunto de treinamento (origem) e o de teste (destino), servindo como dados de meta-treinamento nessa configuração.

Na classificação *few-shot* baseada em reconhecimento visual, a obtenção de uma maneira robusta de extração de características, bem como de medição da similaridade é um fator crucial, especialmente nos casos envolvendo classes de objetos muito semelhantes, com número expres-



sivamente pequeno de imagens disponíveis. Melhorar esses algoritmos é uma medida promissora para reconhecimento de insetos na agricultura, dado que a semelhança entre as espécies torna a sua classificação onerosa. Além disso, a utilização de métodos para diferenciar os insetos benéficos dos prejudiciais à cultura com exatidão, pode ser vista como uma alternativa promissora para o controle biológico de pragas, mitigando, por exemplo, perdas na cultura, bem como a aplicação excessiva de defensivos agrícolas (os agrotóxicos).

O modelo proposto nesse capítulo forneceu dois aprimoramentos relevantes para aumento de desempenho da classificação: a extração e fusão de recursos de baixo e alto níveis de imagens para formar representantes ricos em informações e um classificador baseado em uma medida efetiva de similaridade, a Entropia relativa (RE), em comparação com a distância Euclidiana (ED). Todos os experimentos foram realizados utilizando-se ambas as divergências.

Para a avaliação de desempenho do modelo proposto, foram realizadas comparações relacionadas ao *backbone* e as divergências. Inicialmente, o modelo com extração de recursos globais ( $Backbone^{f4}$ ) foi treinado e testado. Em seguida, ele foi aprimorado com o método FMC para extração e fusão de informações multicamada. A nossa análise considerou tarefas de *1-shot* e *5-shot*. A Tabela 4.3 mostra os resultados da acurácia média usando as duas divergências investigadas. Os dados mostram o desempenho superior do modelo de classificação quando usada a nossa proposta de extração de recursos de imagens FMC.

Tabela 4.3: Acurácia média de cada modelo.

Recurso	<i>1-shot</i>	<i>5-shot</i>
$Backbone^{f4}$ + ED	0,6601	0,7816
$Backbone^{f4}$ + RE	0,6622	0,7882
FMC + ED	0,69	<b>0,8082</b>
FMC + RE	<b>0,6984</b>	0,8014

Conforme os resultados da Tabela 4.3, o  $Backbone^{f4}$ , usando apenas recursos globais da CNN, forneceu os melhores resultados usando a entropia relativa, alcançando 66,22% e 78,82% em *1-shot* e *5-shot* respectivamente. Esses resultados foram melhorados com a fusão de recursos multicamadas usando o extrator FMC. O FMC melhorou o desempenho do modelo em até 3,62% na tarefa de *1-shot*, alcançando 69,84%, e em até 2,66% em *5-shot* com a distância euclidiana, alcançando acurácia de 80,82%. Esses resultados mostram que a fusão de recursos de vários níveis é uma abordagem promissora para melhorar os resultados da classificação *few-shot* de insetos, principalmente usando poucos dados, como em *1-shot*, onde o ganho na acurácia foi maior.

Em concordância com os resultados do Capítulo 3, a entropia relativa demonstrou resultados superiores usando recursos globais ( $Backbone^{f4}$ ) em todas as configurações *K-shot*. No entanto, observou-se que a distância Euclidiana tende a ser melhor em operações de *5-shot* usando os recursos multicamadas, enquanto a entropia relativa tem desempenho inferior. Por outro lado, a entropia supera distância euclidiana em tarefas de *1-shot* usando FMC.

Como demonstrado na Figura 4.7, a acurácia tende a aumentar rapidamente quando os recur-

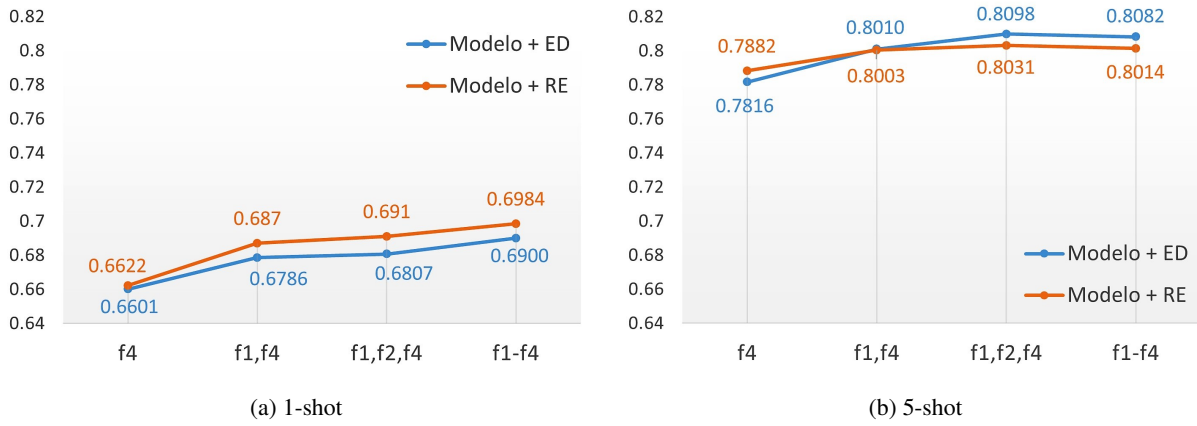


Figura 4.7: Tendência da acurácia a medida que mais recursos  $f$  são fundidos.

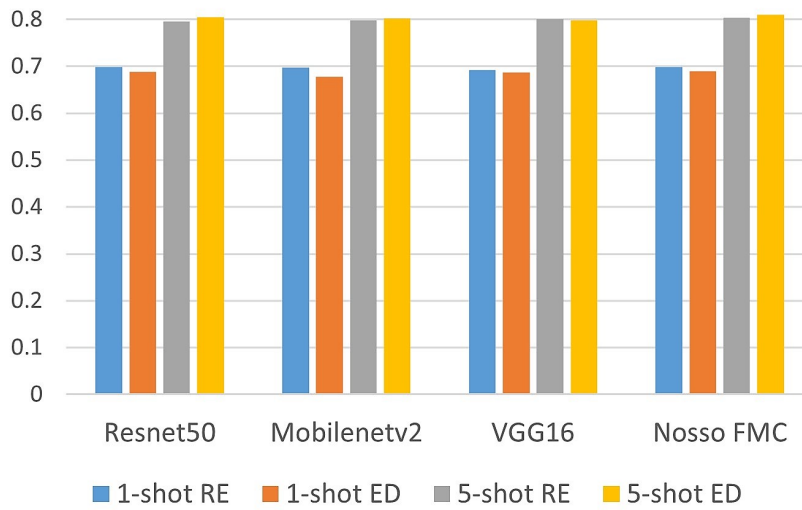


Figura 4.8: Performance of the models.

os locais  $f_1$  e globais  $f_4$  de camadas da CNN são fundidos. A razão disso, é que a função SFR produz *embeddings* mais consistentes em representatividade de classe usando recursos locais e globais, aumentando as chances de acerto durante a classificação. Observou-se que a acurácia aumentou à medida que mais recursos  $f$  de camadas foram fundidos aos recursos globais nas tarefas de *1-shot*. Por outro lado, em *5-shot*, as acurácias reduziram ligeiramente usando todos os recursos  $f$ , com melhor desempenho obtido a partir da fusão dos recursos  $f_1$ ,  $f_2$  e  $f_4$ . Isso demonstra a grande relevância da fusão de recursos multicamada para aplicações de poucos *shots*, como *1-shot*, já que quanto mais informações obtidas sobre classes maior será o aprendizado.

O nosso modelo proposto também foi comparado a outras redes consolidadas da literatura. A Figura 4.8 mostra os desempenhos dos modelos relacionados ao parâmetro  $K$ -shot e a divergência adotada. Os resultados mostram que a distância euclidiana produz melhores acurácias em tarefas de *5-shot* para classificação dos dados de insetos do milho, com exceção da rede VGG16. Por outro lado, a entropia relativa superou a distância euclidiana em todos os *backbones* testados em *1-shot*.

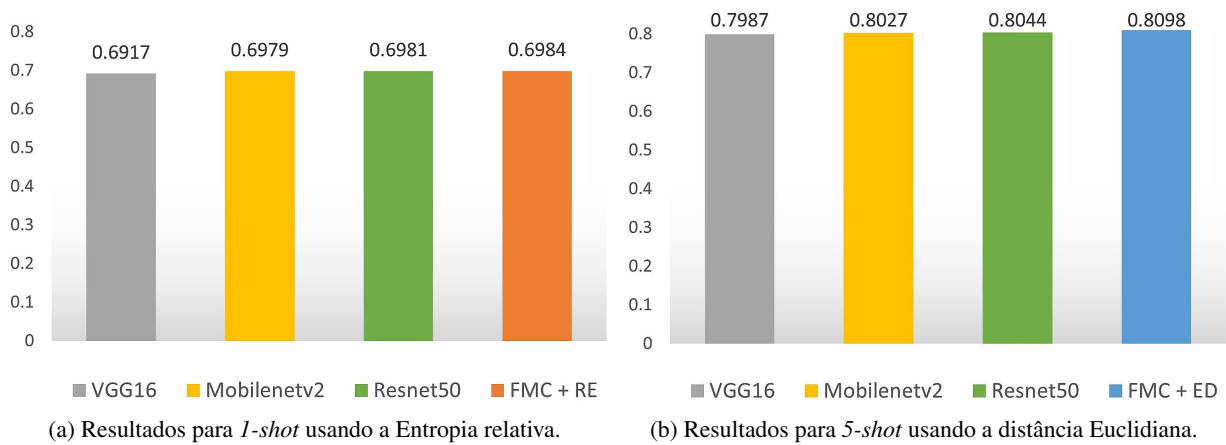


Figura 4.9: Comparação dos melhores desempenhos dos *backbones*.

Projetamos os melhores resultados nos gráficos da Figura 4.9 para uma comparação efetiva entre os desempenhos dos *backbones*. Dessa forma, a Figura 4.9a mostra os resultados de classificação *1-shot* usando a Entropia relativa, e a Figura 4.9b apresenta os resultados de *5-shot* usando a distância Euclidiana.

Como visto na Figura 4.9, o nosso modelo FMC demonstrou o melhor resultado geral entre os *backbones* investigados. Embora esses resultados sejam relativamente aproximados, e podem apresentar pequenas variações, pois a classificação *few-shot* depende da quantidade de tarefas de classificação, existe uma vantagem considerável em nosso modelo relacionada à quantidade de parâmetros de aprendizagem.

Fornecemos um modelo competitivo com redução expressiva de parâmetros de aprendizagem, conforme mostra a Tabela 4.2, apresentado aqui como uma alternativa rápida e eficaz para aplicações de reconhecimento de insetos. O FMC possui cerca de 93% menos parâmetros em relação a Mobilenetv2, e cerca de 99% em relação a ResNet50, os dois *backbones* mais próximos ao FMC em acurácia.

Os resultados gerais mostraram que nosso modelo melhorou a extração tradicional de recursos globais de imagens, fornecendo melhoria da acurácia em até 3,62% em *1-shot* e 2,82% em *5-shot*, considerando as melhores resultados. Com isso, o modelo usando o FMC pode diferenciar pragas e insetos benéficos das culturas de milho com acurácia relevante.

Esses resultados indicam que a fusão de recursos de imagens é fundamentalmente promissor para melhoria dos resultados de classificação *few-shot*. Além disso, a Entropia relativa fornece resultados robustos para medição de similaridade de *embeddings* enriquecidos com informações multicamadas em tarefas de classificação usando uma quantidade muito pequena de dados, com nas tarefas de *1-shot*.

## 4.5 CONCLUSÕES DO CAPÍTULO

Este capítulo apresentou um modelo aprimorado que fornece melhorias para o aprendizado *few-shot* baseado em métrica. O modelo proposto ofereceu duas principais vantagens: a obtenção de um melhor aproveitamento da arquitetura CNN com um mecanismo exclusivo de extração e fusão de recursos visuais multicamada (FMC), e o melhor aproveitamento do classificador baseado na Entropia relativa e a distância Euclidiana. O modelo apresentado foi testado usando um conjunto de dados desafiador de insetos da cultura do milho, com categorias de insetos benéficos e pragas da cultura. O conjunto de dados proposto conta com 28 classes contendo apenas 15 imagens em cada uma delas.

Os resultados mostraram que a fusão de recursos multicamadas melhorou o desempenho do modelo de *few-shot*, com ganho em até 3,62% na acurácia em tarefas de *1-shot*, e 2,82% em *5-shot*, alcançando acurácia geral de 69,84% e 80,82% respectivamente. A Entropia relativa forneceu melhor desempenho na comparação dos *embeddings* gerados pela FMC em tarefas de *1-shot*, e a distância euclidiana, o método mais apropriado em tarefas *5-shot*.

Os resultados foram comparados aos de *backbones* já consolidados na literatura, usados aqui como extratores de recursos de imagens. O nosso modelo forneceu resultados superiores aos alcançados pelos *backbones* ResNet50, VGG16 e Mobilenetv2, com redução de até 99% dos parâmetros de treinamento, e se mostra como um modelo competitivo no cenário de classificação *few-shot*.

# 5 CONCLUSÕES

Esta dissertação forneceu contribuições para a classificação *few-shot* de insetos na agricultura nos seguintes aspectos: forneceu dois conjuntos de dados importantes de insetos da agricultura, analisou as divergências de Bregman como métricas de similaridade em *few-shot* e propôs um modelo aprimorado de aprendizagem. As próximas seções apresentam os detalhes dessas contribuições, sendo também fornecidas algumas sugestões de trabalhos futuros na última seção.

## 5.1 OS CONJUNTOS DE DADOS PROPOSTOS

Este trabalho forneceu dois conjuntos de dados específicos para o reconhecimento automático de insetos na área agrícola. As imagens foram criteriosamente selecionadas para representar, a medida do possível, as situações de capturas reais em campo, contribuindo com a disponibilização de dados para aplicações específicas usando poucas amostras.

O conjunto de dados IP-FSL, apresentado no Capítulo 3, fornece 142 classes de insetos divididas em duas categorias definidas por estágio de maturidade, a fase adulta (97) e a fase primária (45). Ele possui um total de 6817 imagens e pode ser utilizado para outras tarefas de aprendizagem envolvendo visão computacional.

O Capítulo 4 apresentou um conjunto de insetos específico da cultura do milho, abrangendo classes de pragas (17) e de insetos benéficos (11), estes considerados inimigos naturais de pragas do milho. Esses dados têm importância relevante para testes de algoritmos otimizados baseados em poucos dados, fornecendo uma possibilidade inovadora para controle de pragas na agricultura.

## 5.2 AS DIVERGÊNCIAS DE BREGMAN EM FSL

O capítulo 3 forneceu, provavelmente, o primeiro estudo sobre a comparação das divergências de Bregman para classificação de insetos usando aprendizagem *few-shot*. Os resultados mostraram que mesmo em tarefas de classificação com grau de similaridade elevado entre classes, os modelos baseados nas divergências assimétricas produziram medidas de similaridade mais promissoras para classificação de insetos usando imagens de campo, mostrando serem soluções tão competitivas quanto as métricas simétricas utilizadas com frequência em FSL.

Além da análise de divergências, o estudo apontou caminhos para a escolha dos parâmetros *N-way-K-shot*, em que foram demonstradas as melhores condições de aprendizado dos modelos, bem como os pontos com menores possibilidades de erro de classificação. O estudo forneceu uma

importante base para futuras pesquisas envolvendo o aprendizado *few-shot* baseado em métrica e classificação de insetos na agricultura, com importante atenção no ciclo de vida do inseto, o que foi pouco explorado na literatura.

### 5.3 MODELO DE CLASSIFICAÇÃO PROPOSTO

A extração de *embeddings* ricos em informações de imagens na aprendizagem *few-shot* é uma prática fundamental, dada a pequena quantidade de dados disponível para aprendizagem. Esta pesquisa forneceu um modelo aprimorado de FSL baseado em um mecanismo que combina dois elementos: extração e fusão de recursos multicamadas da CNN (FMC), e a classificação baseada no melhor desempenho entre a Entropia relativa e a distância Euclidiana.

O modelo proposto faz um melhor uso da CNN para construção de *embeddings* a partir da fusão de recursos, melhorando a acurácia em até 3,62% nas tarefas de classificação de *1-shot* e 2,82% em tarefas de *5-shot* em relação a alcançada usando um *backbone* tradicional, com extração de recursos globais de imagem. Além disso, o modelo forneceu resultados superiores aos obtidos com o uso das redes ResNet50, VGG16 e Mobilenetv2, usadas aqui como *backbones* para extração de recursos de imagens, com redução de parâmetros de aprendizagem em até 99%.

Além de fornecer um modelo aprimorado de FSL, essa pesquisa apresentou um estudo inovador, com potencial de aplicação no controle biológico de pragas em culturas de milho, combinando uma abordagem de aprendizagem *few-shot* a uma tarefa desafiadora de reconhecimento visual de insetos, a qual demanda soluções de classificação rápida e com acurácia relevante para aplicação em campo.

### 5.4 SUGESTÕES DE TRABALHOS FUTUROS

A aprendizagem profunda tem sido uma boa escolha para soluções automáticas de problemas agrícolas. No entanto, a necessidade de *hardwares* potentes e de grandes conjuntos de dados para treinamento dos modelos dificulta a ampla disseminação dessa tecnologia, limitando a aplicação do reconhecimento automático ao nível de laboratório em alguns casos. Como trabalho futuro, pretende-se realizar a otimização do modelo proposto nesse trabalho para aplicações embarcadas de baixo custo. A combinação de dispositivos reconfiguráveis, como FPGA, processadores ARM e *softwares* livres como o linux, pode ser uma solução relevante para o desenvolvimento de sistemas embarcados, incluindo o uso de drones. Essa tecnologia pode favorecer a disseminação de sistemas de reconhecimento rápido e acessível para uso no campo, contribuindo com agricultores que possuem recursos financeiros incompatíveis com altos investimentos e aplicações remotas.

## 5.5 ARTIGOS PUBLICADOS

Como resultado dessa pesquisa, um trabalho voltado para aplicação foi publicado na revista *Agronomy*, a qual possui Fator de Impacto de 3,949 (2021) e avaliação Qualis referente à A2, conforme citação abaixo:

GOMES, Jacó C.; BORGES, Dábio L. Insect Pest Image Recognition: A Few-Shot Machine Learning Approach including Maturity Stages Classification. ***Agronomy***, v. 12, n. 8, p. 1733, 2022.

# REFERÊNCIAS BIBLIOGRÁFICAS

- ALVES, A. N.; SOUZA, W. S.; BORGES, D. L. Cotton pests classification in field-based images using deep residual networks. *Computers and Electronics in Agriculture*, Elsevier, v. 174, p. 105488, 2020.
- AMIN, S. U.; ALSULAIMAN, M.; MUHAMMAD, G.; BENCHERIF, M. A.; HOSSAIN, M. S. Multilevel weighted feature fusion using convolutional neural networks for eeg motor imagery classification. *IEEE Access*, IEEE, v. 7, p. 18940–18950, 2019.
- ARGÜESO, D.; PICON, A.; IRUSTA, U.; MEDELA, A.; SAN-EMETERIO, M. G.; BERECIARTUA, A.; ALVAREZ-GILA, A. Few-Shot Learning approach for plant disease classification using images taken in the field. *Computers and Electronics in Agriculture*, Elsevier, v. 175, p. 105542, 2020. ISSN 0168-1699.
- BANERJEE, A.; MERUGU, S.; DHILLON, I. S.; GHOSH, J.; LAFFERTY, J. Clustering with bregman divergences. *Journal of machine learning research*, v. 6, n. 10, 2005.
- BATENI, P.; GOYAL, R.; MASRANI, V.; WOOD, F.; SIGAL, L. Improved few-shot visual classification. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington: IEEE, 2020. p. 14493–14502.
- BRÉCHETEAU, C.; FISCHER, A.; LEVRARD, C. Robust bregman clustering. *The Annals of Statistics*, Institute of Mathematical Statistics, v. 49, n. 3, p. 1679–1701, 2021.
- CHEN, Y.; LIU, Z.; XU, H.; DARRELL, T.; WANG, X. Meta-baseline: Exploring simple meta-learning for few-shot learning. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. [S.l.: s.n.], 2021. p. 9062–9071.
- CILINGIR, H. K.; MANZELLI, R.; KULIS, B. Deep divergence learning. In: *Proceedings of the 37th International Conference on Machine Learning*. PMLR, 2020. (Proceedings of Machine Learning Research, v. 119), p. 2027–2037. Disponível em: <https://proceedings.mlr.press/v119/cilingir20a.html>.
- DHILLON, G. S.; CHAUDHARI, P.; RAVICHANDRAN, A.; SOATTO, S. A Baseline for Few-Shot Image Classification. In: *International Conference on Learning Representations*. New Orleans: ICRL, 2020. p. 20.
- DING, Y.; TIAN, X.; YIN, L.; CHEN, X.; LIU, S.; YANG, B.; ZHENG, W. Multi-scale relation network for few-shot learning based on meta-learning. In: SPRINGER. *International Conference on Computer Vision Systems*. [S.l.], 2019. p. 343–352.
- ENDRES, D. M.; SCHINDELIN, J. E. A new metric for probability distributions. *IEEE Transactions on Information theory*, IEEE, v. 49, n. 7, p. 1858–1860, 2003.
- ES-SAADY, Y.; MASSI, I. E.; YASSA, M. E.; MAMMASS, D.; BENAZOUN, A. Automatic recognition of plant leaves diseases based on serial combination of two svm classifiers. In: IEEE. *2016 International Conference on Electrical and Information Technologies (ICEIT)*. 2016. p. 561–566. Disponível em: <https://ieeexplore.ieee.org/abstract/document/7519661>.
- FE-FEI, L. et al. A bayesian approach to unsupervised one-shot learning of object categories. In: IEEE. *proceedings ninth IEEE international conference on computer vision*. [S.l.], 2003. p. 1134–1141.
- FU, X.; ZHAO, Y.; WEI, Y.; ZHAO, Y.; WEI, S. Rich features embedding for cross-modal retrieval: A simple baseline. *IEEE Transactions on Multimedia*, IEEE, v. 22, n. 9, p. 2354–2365, 2019.



- GOMES, J. C.; BORGES, D. L. Insect pest image recognition: A few-shot machine learning approach including maturity stages classification. *Agronomy*, Multidisciplinary Digital Publishing Institute, v. 12, n. 8, p. 1733, 2022.
- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2016. p. 770–778.
- HOCHREITER, S.; YOUNGER, A. S.; CONWELL, P. R. Learning to learn using gradient descent. In: *International Conference on Artificial Neural Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001. p. 87–94.
- HOFFER, E.; AILON, N. Deep metric learning using triplet network. In: *International workshop on similarity-based pattern recognition*. [S.l.]: Springer, 2015. p. 84–92.
- ISLAM, M.; DINH, A.; WAHID, K.; BHOWMIK, P. Detection of potato diseases using image segmentation and multiclass support vector machine. In: IEEE. *2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE)*. 2017. p. 1–4. Disponível em: <https://ieeexplore.ieee.org/abstract/document/7946594>).
- JAMAL, M. A.; QI, G.-J. Task agnostic meta-learning for few-shot learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2019. p. 11719–11727.
- JI, Z.; CHAI, X.; YU, Y.; PANG, Y.; ZHANG, Z. Improved prototypical networks for few-shot learning. *Pattern Recognition Letters*, Elsevier, v. 140, p. 81–87, 2020.
- JIANG, W.; HUANG, K.; GENG, J.; DENG, X. Multi-scale metric learning for few-shot learning. *IEEE Transactions on Circuits and Systems for Video Technology*, IEEE, v. 31, n. 3, p. 1091–1102, 2020.
- KANG, B.; LIU, Z.; WANG, X.; YU, F.; FENG, J.; DARRELL, T. Few-shot object detection via feature reweighting. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul, Korea: ICCV, 2019. p. 8420–8429.
- KASINATHAN, T.; SINGARAJU, D.; UYYALA, S. R. Insect classification and detection in field crops using modern machine learning techniques. *Information Processing in Agriculture*, Elsevier, v. 8, n. 3, p. 446–457, 2021.
- KAYA, M.; BILGE, H. Ş. Deep metric learning: A survey. *Symmetry*, Multidisciplinary Digital Publishing Institute, v. 11, n. 9, p. 1066, 2019.
- KIM, J.; CHI, M. Saffnet: Self-attention-based feature fusion network for remote sensing few-shot scene classification. *Remote Sensing*, MDPI, v. 13, n. 13, p. 2532, 2021.
- KOCH, G.; ZEMEL, R.; SALAKHUTDINOV, R. et al. Siamese neural networks for one-shot image recognition. In: LILLE. *ICML deep learning workshop*. [S.l.], 2015. v. 2, p. 0.
- KUSRINI, K.; SUPUTA, S.; SETYANTO, A.; AGASTYA, I. M. A.; PRIANTORO, H.; CHANDRAMOULI, K.; IZQUIERDO, E. Data augmentation for automated pest classification in mango farms. *Computers and Electronics in Agriculture*, Elsevier, v. 179, p. 105842, 2020.
- LI, Y.; CHAO, X. Ann-based continual classification in agriculture. *Agriculture*, Multidisciplinary Digital Publishing Institute, v. 10, n. 5, p. 178, 2020.
- LI, Y.; CHAO, X. Semi-supervised few-shot learning approach for plant diseases recognition. *Plant Methods*, Springer, v. 17, n. 1, p. 1–10, 2021.
- LI, Y.; NIE, J.; CHAO, X. Do we really need deep cnn for plant diseases identification? *Computers and Electronics in Agriculture*, Elsevier, v. 178, p. 105803, 2020.

- LI, Y.; WANG, H.; DANG, L. M.; SADEGHI-NIARAKI, A.; MOON, H. Crop pest recognition in natural scenes using convolutional neural networks. *Computers and Electronics in Agriculture*, Elsevier, v. 169, p. 105174, 2020.
- LI, Y.; YANG, J. Few-shot cotton pest recognition and terminal realization. *Computers and Electronics in Agriculture*, Elsevier, v. 169, p. 105240, 2020.
- LI, Y.; YANG, J. Meta-learning baselines and database for few-shot classification in agriculture. *Computers and Electronics in Agriculture*, Elsevier, v. 182, p. 106055, 2021.
- LIM, J. Y.; LIM, K. M.; OOI, S. Y.; LEE, C. P. Efficient-prototypicalnet with self knowledge distillation for few-shot learning. *Neurocomputing*, Elsevier, v. 459, p. 327–337, 2021.
- LIN, C.; GAO, F. An extension of prototypical networks. In: IEEE. *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*. [S.l.], 2020. v. 1, p. 421–425.
- LIN, T.-Y.; DOLLÁR, P.; GIRSHICK, R.; HE, K.; HARIHARAN, B.; BELONGIE, S. Feature pyramid networks for object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2017. p. 2117–2125.
- LIU, J.; WANG, X. Tomato diseases and pests detection based on improved yolo v3 convolutional neural network. *Frontiers in plant science*, Frontiers, v. 11, p. 898, 2020.
- LIU, X.; ZHOU, F.; LIU, J.; JIANG, L. Meta-learning based prototype-relation network for few-shot classification. *Neurocomputing*, Elsevier, v. 383, p. 224–234, 2020.
- LU, J.; GONG, P.; YE, J.; ZHANG, C. Learning from very few samples: A survey. *arXiv preprint arXiv:2009.02653*, 2020.
- MILLER, E. G.; MATSAKIS, N. E.; VIOLA, P. A. Learning from one example through shared densities on transforms. In: IEEE. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662)*. [S.l.], 2000. v. 1, p. 464–471.
- NGUGI, L. C.; ABELWAHAB, M.; ABO-ZAHHAD, M. Recent advances in image processing techniques for automated leaf pest and disease recognition—a review. *Information processing in agriculture*, Elsevier, v. 8, n. 1, p. 27–51, 2021.
- NUTHALAPATI, S. V.; TUNGA, A. Multi-domain few-shot learning and dataset for agricultural applications. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision. ICCV*, 2021. p. 1399–1408. Disponível em: <https://iccv2021.thecvf.com/>.
- OLIVEIRA, C.; AUAD, A.; MENDES, S.; FRIZZAS, M. Crop losses and the economic impact of insect pests on brazilian agriculture. *Crop Protection*, Elsevier, v. 56, p. 50–54, 2014.
- ORESHKIN, B.; LÓPEZ, P. R.; LACOSTE, A. Tadam: Task dependent adaptive metric for improved few-shot learning. *Advances in neural information processing systems*, v. 31, 2018.
- PENG, B.; ZHU, C.; LI, C.; LI, X.; LI, J.; ZENG, M.; GAO, J. Few-shot Natural Language Generation for Task-Oriented Dialog. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. [S.l.: s.n.], 2020. p. 172–182.
- PINTO, A. de S.; PARRA, J. R. P.; OLIVEIRA, H. N. de. *Guia ilustrado de pragas e insetos benéficos do milho e sorgo*. Ribeirão Preto: AS Pinto, 2004.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATY, A.; KHOSLA, A.; BERNSTEIN, M. et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, Springer, v. 115, n. 3, p. 211–252, 2015.

SANDLER, M.; HOWARD, A.; ZHU, M.; ZHMOGINOV, A.; CHEN, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 4510–4520.

SIAHKAMARI, A.; XIA, X.; SALIGRAMA, V.; CASTAÑÓN, D.; KULIS, B. Learning to approximate a bregman divergence. *Advances in Neural Information Processing Systems*, v. 33, p. 3603–3612, 2020.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. In: *ICLR*. [S.l.: s.n.], 2015.

SNELL, J.; SWERSKY, K.; ZEMEL, R. Prototypical networks for few-shot learning. In: *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. v. 30. Disponível em: <https://proceedings.neurips.cc/paper/2017/file/cb8da6767461f2812ae4290eac7cbc42-Paper.pdf>.

SUN, Q.; LIU, Y.; CHUA, T.-S.; SCHIELE, B. Meta-transfer learning for few-shot learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2019. p. 403–412.

SUNG, F.; YANG, Y.; ZHANG, L.; XIANG, T.; TORR, P. H. S.; HOSPEDALES, T. M. Learning to compare: Relation network for few-shot learning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. [S.l.: s.n.], 2018. p. 1199–1208.

THENMOZHI, K.; REDDY, U. S. Crop pest classification based on deep convolutional neural network and transfer learning. *Computers and Electronics in Agriculture*, Elsevier, v. 164, p. 104906, 2019.

VINYALS, O.; BLUNDELL, C.; LILLICRAP, T.; WIERSTRA, D. et al. Matching networks for one shot learning. *Advances in neural information processing systems*, v. 29, p. 3630–3638, 2016.

WANG, Y.; WANG, S. Imal: An improved meta-learning approach for few-shot classification of plant diseases. In: IEEE. *2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE)*. [S.l.], 2021. p. 1–7.

WANG, Y.; YAO, Q.; KWOK, J. T.; NI, L. M. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, ACM New York, NY, USA, v. 53, n. 3, p. 1–34, 2020.

WU, X.; ZHAN, C.; LAI, Y.-K.; CHENG, M.-M.; YANG, J. Ip102: A large-scale benchmark dataset for insect pest recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, CA: CVPR, 2019. p. 8787–8796.

YANG, J.; GUO, X.; LI, Y.; MARINELLO, F.; ERCISLI, S.; ZHANG, Z. A survey of few-shot learning in smart agriculture: developments, applications, and challenges. *Plant Methods*, BioMed Central, v. 18, n. 1, p. 1–12, 2022.

YANG, S.; LIU, F.; DONG, N.; WU, J. Comparative analysis on classical meta-metric models for few-shot learning. *IEEE Access*, IEEE, v. 8, p. 127065–127073, 2020.

YANG, Z.; YANG, X.; LI, M.; LI, W. Small-sample learning with salient-region detection and center neighbor loss for insect recognition in real-world complex scenarios. *Computers and Electronics in Agriculture*, Elsevier, v. 185, p. 106122, 2021. ISSN 0168-1699.

YE, H.-J.; HU, H.; ZHAN, D.-C.; SHA, F. Few-shot learning via embedding adaptation with set-to-set functions. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2020. p. 8808–8817.

ZEILER, M. D.; FERGUS, R. Visualizing and understanding convolutional networks. In: SPRINGER. *European conference on computer vision*. Zurich, Switzerland, 2014. p. 818–833.

ZHANG, X.; QIAO, Y.; MENG, F.; FAN, C.; ZHANG, M. Identification of maize leaf diseases using improved deep convolutional neural networks. *Ieee Access*, IEEE, v. 6, p. 30370–30377, 2018.

# APÊNDICES

## A. INFORMAÇÕES DO CONJUNTO DE DADOS IP-FSL

Tabela A.1: Informações dos dados IP-FSL. Essa é uma versão compacta do conjunto de dados IP102 (WU et al., 2019) construída especificamente para aprendizagem *few-shot*. Foram mantidos os nomes dos insetos conforme publicado no artigo de referência, podendo conter nomes comuns, bem como científicos.

Nome	Fase adulta	Fase jovem	Nome	Fase adulta	Fase jovem
1 <i>rice leaf roller</i>	50	50	52 <i>blister beetle</i>	50	-
2 <i>rice leaf caterpillar</i>	50	50	53 <i>therioaphis maculata</i> Buckton	50	-
3 <i>paddy stem maggot</i>	50	50	54 <i>odonthrips loti</i>	50	-
4 <i>asiatic rice borer</i>	50	50	55 <i>Thrips</i>	50	-
5 <i>yellow rice borer</i>	50	50	56 <i>alfalfa seed chalcid</i>	50	-
6 <i>rice gall midge</i>	50	31	57 <i>Pieris canidia</i>	50	-
7 <i>Rice Stemfly</i>	50	47	58 <i>Apolygus lucorum</i>	50	-
8 <i>brown plant hopper</i>	50	17	59 <i>Limacodidae</i>	50	50
9 <i>white backed plant hopper</i>	50	18	60 <i>Viteus vitifoliae</i>	-	50
10 <i>small brown plant hopper</i>	50	-	61 <i>Colomerus vitis</i>	-	50
11 <i>rice water weevil</i>	50	50	62 <i>Brevipoalpus lewisi</i> McGregor	47	-
12 <i>rice leafhopper</i>	50	-	63 <i>oides decempunctata</i>	50	-
13 <i>grain spreader thrips</i>	50	-	64 <i>Polyphagotars onemus latus</i>	50	-
14 <i>rice shell pest</i>	50	50	65 <i>Pseudococcus comstocki</i> Kuwana	50	-
15 <i>grub</i>	-	50	66 <i>parathrene regalis</i>	40	30
16 <i>mole cricket</i>	50	-	67 <i>Ampelophaga</i>	50	50
17 <i>wireworm</i>	50	50	68 <i>Lycorma delicatula</i>	50	-
18 <i>white margined moth</i>	26	50	69 <i>Xylotrechus</i>	50	-
19 <i>black cutworm</i>	50	50	70 <i>Cicadella viridis</i>	50	-
20 <i>large cutworm</i>	50	50	71 <i>Miridae</i>	50	-
21 <i>yellow cutworm</i>	50	50	72 <i>Trialeurodes vaporariorum</i>	50	-
22 <i>red spider</i>	50	-	73 <i>Erythroneura apicalis</i>	42	-
23 <i>corn borer</i>	50	50	74 <i>Papilio xuthus</i>	50	50
24 <i>army worm</i>	35	50	75 <i>Panonychus citri</i> McGregor	50	-
25 <i>aphids</i>	50	-	76 <i>Phyllocoptes oleiverus</i> ashmead	-	50
26 <i>Potosiabre vitarsis</i>	50	-	77 <i>Icerya purchasi</i> Maskell	50	-
27 <i>peach borer</i>	50	50	78 <i>Unaspis yanonensis</i>	50	-
28 <i>english grain aphid</i>	50	-	79 <i>Ceroplastes rubens</i>	50	-
29 <i>green bug</i>	50	-	80 <i>Chrysomphalus aonidium</i>	50	-
30 <i>bird cherry-oataphid</i>	50	-	81 <i>Parlatoria zizyphus</i> Lucas	44	-
31 <i>wheat blossom midge</i>	50	50	82 <i>Nipaeococcus vastalor</i>	50	-
32 <i>penthaleus major</i>	50	-	83 <i>Aleurocanthus spiniferus</i>	-	50
33 <i>longlegged spider mite</i>	50	-	84 <i>Tetradacus c Bactrocera minax</i>	50	50
34 <i>wheat phloeothrips</i>	50	-	85 <i>Dacus dorsalis</i> (Hendel)	50	40
35 <i>wheat sawfly</i>	50	50	86 <i>Bactrocera tsuneonis</i>	50	20
36 <i>cerodonta denticornis</i>	50	32	87 <i>Prodenia litura</i>	50	50
37 <i>beet fly</i>	50	-	88 <i>Adristyrannus</i>	50	40
38 <i>lea beetle</i>	50	-	89 <i>Phyllocnistis citrella</i> Stainton	50	50
39 <i>cabbage army worm</i>	50	50	90 <i>Toxoptera citricidus</i>	50	-
40 <i>beet army worm</i>	50	50	91 <i>Toxoptera aurantii</i>	50	-
41 <i>Beet spot flies</i>	50	50	92 <i>Aphis citricola</i> Vander Goot	50	-
42 <i>meadow moth</i>	50	25	93 <i>Scirtothrips dorsalis</i> Hood	50	-
43 <i>beet weevil</i>	50	-	94 <i>Dasineura</i> sp	33	50
44 <i>sericaorient alismots chulsky</i>	50	-	95 <i>Lawana imitata</i> Melichar	50	-
45 <i>alfalfa weevil</i>	50	50	96 <i>Salurnis marginella</i> Guerr	50	-
46 <i>flax budworm</i>	50	50	97 <i>Deporaus marginatus</i> Pascoe	50	-
47 <i>alfalfa plant bug</i>	50	-	98 <i>Chlumetia transversa</i>	50	50
48 <i>tarnished plant bug</i>	50	-	99 <i>Mango flat beak leafhopper</i>	50	-
49 <i>Locustoidea</i>	50	-	100 <i>Rhytidodera bowrinii</i> white	50	-
50 <i>lytta polita</i>	50	-	101 <i>Sternochetus frigidus</i>	50	-
51 <i>legume blister beetle</i>	50	-	102 <i>Cicadellidae</i>	50	-
<b>Total de amostras</b>	<b>2461</b>	<b>1320</b>	<b>Total de amostras</b>	<b>2306</b>	<b>730</b>

## B. RESUMO DO MODELO PROPOSTO PARA EXTRAÇÃO DE RECURSOS MULTICAMADA

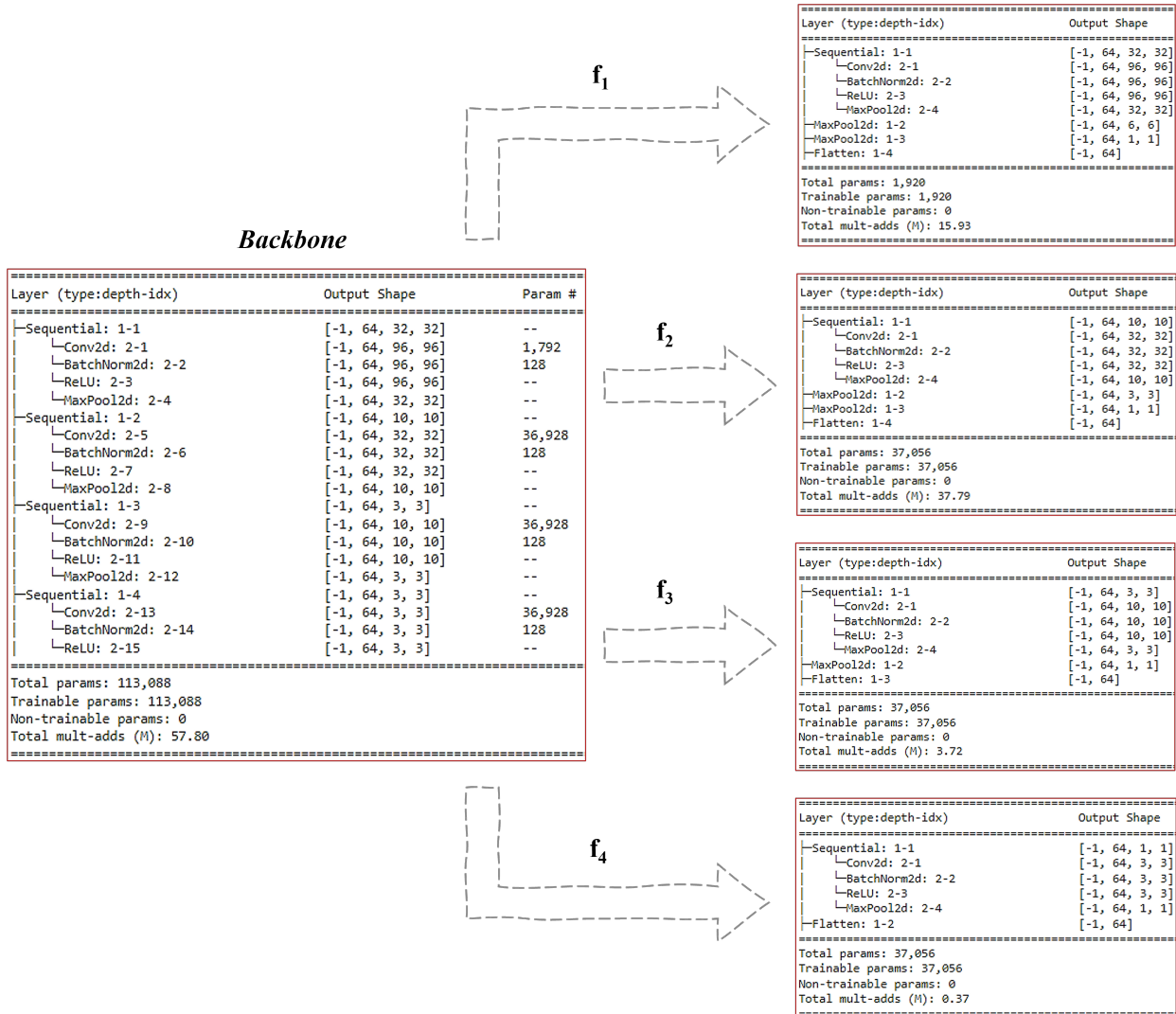


Figura B.1: Resumo do modelo de extração de recursos multicamada, composto pelo *backbone* principal e suas ramificações  $f_x$  ( $x = 1, 2, 3, 4$ ).