UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE COMPUTAÇÃO

Exame de Qualificação de Mestrado

10 de Dezembro de 2020

Aumentação de Dados para Classificação de Imagens Utilizando Redes Neurais Profundas

Candidata: Marianna de Pinho Severo

Orientador: Prof. Dr. Zanoni Dias

Coorientador: Prof. Dr. Hélio Pedrini

Sumário

1 Introdução **1**
- 1.1 Caracterização do Problema . . . 1
- 1.2 Objetivos e Contribuições . . . 3
- 1.3 Questões de Pesquisa . . . 4
- 1.4 Organização do Texto . . . 4

2 Revisão Bibliográfica **5**
- 2.1 Métodos Básicos de Aumentação de Dados . . . 5
- 2.2 Aumentação Baseada em Aprendizado Profundo . . . 9
  - 2.2.1 Aumentação de Dados no Espaço de Características . . . 9
  - 2.2.2 Aumentação Baseada em Redes Adversariais Generativas . . . 10
  - 2.2.3 Aumentação Baseada em Transferência de Estilo Neural . . . 13
  - 2.2.4 Aumentação de Dados Baseada em Meta-Aprendizado . . . 16
- 2.3 Resumo dos Trabalhos Relacionados . . . 18

3 Materiais **21**
- 3.1 Bases de Dados . . . 21
  - 3.1.1 VGGFace2 . . . 21
  - 3.1.2 Places365 - Standard . . . 22
  - 3.1.3 Digipathos Embrapa . . . 24
  - 3.1.4 COVIDx . . . 24
  - 3.1.5 Resumo das Bases . . . 24
- 3.2 Recursos Computacionais . . . 24

4 Métodos **28**
- 4.1 Metodologia . . . 28
  - 4.1.1 Treinamento, Validação e Teste . . . 29
  - 4.1.2 Otimização de Hiperparâmetros e Número de Épocas . . . 29
  - 4.1.3 Protocolo Experimental . . . 30
- 4.2 Métricas de Avaliação de Desempenho . . . 31
  - 4.2.1 Sensibilidade . . . 31
  - 4.2.2 Acurácia . . . 32
  - 4.2.3 Precisão . . . 33
  - 4.2.4 Outras Métricas . . . 33

5 Resultados Preliminares **35**
5.1 Resultados e Discussões 36

6 Plano de Trabalho e Cronograma de Execução **42**

Bibliografia **44**

Resumo

Nos últimos anos, com o incremento do poder computacional, o aumento da disponibilidade de dados (*Big Data*) e o desenvolvimento de algoritmos cada vez mais sofisticados, as pesquisas em Aprendizado de Máquina Profundo têm trazido grandes transformações para diversos setores da sociedade. Exemplos são a construção de sistemas de diagnóstico médico, de agricultura inteligente, de segurança, de auxílio na educação e de criação de ambientes mais inclusivos. Apesar disso, a escassez de dados em determinados domínios de aplicação ainda representa um dos principais problemas enfrentados por esse campo de pesquisas, desencadeando desafios relacionados ao desempenho e à justiça dos algoritmos. Neste projeto, propomos a investigação de técnicas de aumentação de dados para a adição de volume e diversidade em conjuntos de dados destinados a tarefas de classificação de imagens utilizando Aprendizado de Máquina Profundo. Para isso, diferentes abordagens de aumentação de dados serão analisadas e combinadas, experimentaremos sua capacidade de generalização para bases de dados pertencentes a variados domínios de aplicação e níveis de dificuldade, e as compararemos com outras abordagens que lidam com os problemas decorrentes da escassez de dados. Como resultados iniciais, treinamos redes neurais convolucionais para a classificação de imagens de radiografia de tórax, mostrando que o emprego de técnicas básicas de aumentação de dados contribuiu para o melhoramento dos modelos.

Capítulo 1

Introdução

Visão Computacional é um campo de pesquisa que tem proporcionado a melhoria de atividades realizadas em diversos setores da vida em sociedade, como aquelas desempenhadas na indústria, na agricultura e na área médica. Um dos grandes avanços que vêm impulsionando fortemente o desenvolvimento desse campo é a criação das técnicas de Aprendizado Profundo (*Deep Learning*), especialmente aquelas que englobam as Redes Neurais Convolucionais (*Convolutional Neural Networks*) [72].

Entretanto, os métodos de aprendizado profundo possuem uma característica que os torna de grande potencial para a melhoria das tarefas de visão computacional, ao mesmo tempo que traz grandes dificuldades para sua utilização em alguns domínios de aplicação: o desempenho desses métodos melhora conforme aumenta-se o tamanho e a diversidade do conjunto de dados de treinamento, ao passo que ele diminui à medida que se reduz essas características [74]. Ou seja, o Aprendizado Profundo depende de grandes quantidades de dados para que possa gerar bons resultados, quando comparados com aqueles obtidos por agentes humanos ou por outras técnicas de visão computacional no mesmo domínio de aplicação.

Tendo em vista que, em inúmeras aplicações, não se consegue ter acesso a grandes volumes de dados de treinamento – seja por questões de custo, de dificuldades em obtê-los ou em rotulá-los –, diversos métodos têm sido desenvolvidos para tratar esse problema, alguns deles tendo como foco as arquiteturas de redes neurais a serem empregadas e outros tratando do próprio conjunto de dados [72].

Neste capítulo, descreveremos o problema a ser abordado durante este projeto, os objetivos e as contribuições esperadas, as questões de pesquisa que guiarão o seu desenvolvimento e a organização do restante do texto.

1.1 Caracterização do Problema

A computação tem avançado bastante ao longo dos anos, assim como outras áreas, tais como a eletrônica e as tecnologias de comunicação. Isso tem permitido que novas tecnologias insiram-se no cotidiano de um número cada vez maior de pessoas, auxiliando em suas atividades, tornando mais eficientes algumas delas e proporcionando avanços em diferentes áreas de conhecimento.

Um dos campos de estudos dentro da Computação que têm ganhado cada vez mais foco e se beneficiado desses avanços é o de Visão Computacional. Este tem como objetivo extrair informações de imagens de maneira que se possa fazer inferências sobre o mundo real [80]. Em outras palavras, a Visão Computacional visa criar sistemas computacionais que sejam capazes de realizar as tarefas que o sistema visual humano consegue, contribuindo para tornar as atividades que decorrem dessas capacidades mais eficientes, baratas e eficazes.

Dentre as inúmeras aplicações que esse campo de pesquisas possui, podemos enumerar algumas, como:

- A identificação de indivíduos doentes e espécies indesejadas em plantações, o que permite o tratamento precoce e a consequente diminuição de perda de colheitas, sendo uma atividade de grande importância tendo em vista o papel fundamental que a produção de alimentos possui para nossa subsistência [5, 37];
- A inspeção de máquinas em indústrias, de maneira a prevenir acidentes e evitar perdas de equipamentos e de produção [80];
- A análise de imagens médicas, que tem permitido a realização de diagnósticos mais rápidos e precisos, e a identificação de problemas de saúde mais difíceis de serem detectados, contribuindo para a manutenção de inúmeras vidas [80, 83];
- Atividades de monitoramento, que podem envolver desde tarefas ligadas à segurança, como o reconhecimento de impressões digitais, a identificação de faces e o reconhecimento de placas de carros, até aquelas relacionadas ao estudo da cobertura de terrenos modificados por ações humanas ou fenômenos naturais [27, 80];
- Tarefas que envolvem a vigilância e o cuidado de pessoas em vulnerabilidade, tais como o reconhecimento de atividades em casas inteligentes [40].

Nesse contexto, uma das áreas que têm se mostrado muito promissora para ajudar na melhoria do campo de Visão Computacional é a de Aprendizado Profundo, a qual tem apresentado os melhores resultados quando empregada em tarefas de aprendizado supervisionado. Além disso, o Aprendizado Profundo vem sendo o foco de inúmeros estudos, não somente relacionados à visão computacional, que mostram que o desempenho de seus algoritmos melhora significantemente conforme aumenta-se o tamanho do conjunto de dados de treinamento [72, 74].

Entretanto, em diversos domínios de aplicação não se tem acesso a grandes volumes de dados, como é o caso da análise de imagens médicas, onde encontra-se grande dificuldade para a geração de bases de dados, devido à necessidade de especialistas para a rotulação das imagens, ao alto custo para sua geração e à raridade de algumas doenças [81]. Além do tamanho do conjunto de dados de treinamento, outro fator que influencia o desempenho dos modelos de aprendizado que utilizam determinado conjunto é a diversidade dos dados presentes nele. No mundo real, um mesmo objeto ou cena podem ser vistos sob diferentes perspectivas, condições de luminosidade, escala, plano de fundo e situações de obstrução. Assim, um conjunto de dados que possui imagens que englobam essas características, ou

seja, que é diverso, pode ajudar na produção de resultados muito melhores do que aqueles obtidos quando essa variedade não está presente [72].

Portanto, o emprego de conjuntos de dados pequenos e com pouca diversidade para o treinamento de modelos de aprendizado profundo gera um problema chamado de *overfitting*, em que os modelos operam muito bem com os dados de treinamento, mas apresentam péssimos resultados para os dados de teste. Isso tem se apresentado como um dos grandes desafios enfrentados pelo campo de Aprendizado Profundo e inúmeros trabalhos têm sido realizados para o desenvolvimento de modelos de aprendizado que tenham grande capacidade de generalização, ou seja, que tenham resultados tão bons para dados ainda não vistos - como os dados de teste e os encontrados em operação no mundo real - como aqueles obtidos durante o treinamento, prevenindo ou reduzindo a ocorrência de *overfitting* [72].

Dada a complexidade desse problema, diversos trabalhos têm sido desenvolvidos na literatura, alguns deles tendo em foco as arquiteturas dos modelos de aprendizado [33, 42]; outros desenvolvendo técnicas que alteram parâmetros dos modelos e dos espaços de características, tais como regularização, normalização, pré-treino e transferência de aprendizado [19, 35, 38, 87]; e outros que têm como foco os conjuntos de dados em si, nos quais se destacam técnicas de aumentação de dados [12, 21, 45, 62].

1.2 Objetivos e Contribuições

Os principais objetivos deste trabalho são, mas não estão limitados, a realização de uma análise comparativa sobre um conjunto diverso de métodos de aumentação de dados do estado da arte. Também, dado o grande potencial da combinação desses métodos para a melhoria dos modelos de aprendizado profundo [55], a investigação das vantagens e desvantagens de sua aplicação individual e conjunta, além da análise da capacidade de generalização dessas abordagens para a classificação de imagens em bases de dados pertencentes a diferentes domínios de aplicação e níveis de dificuldade. Para atingir esse propósito, alguns objetivos específicos precisam ser alcançados:

- Estudar as abordagens do estado da arte utilizadas para classificação de imagens;
- Estudar técnicas do estado da arte para a aumentação de dados, pertencentes a variadas abordagens de aumentação (tradicional, do espaço de características, entre outras);
- Escolher e avaliar os métodos de aumentação do estado da arte em bases de dados voltadas para diferentes domínios de aplicação;
- Investigar os benefícios e desvantagens de diversas combinações das técnicas escolhidas, assim como sua capacidade de generalização entre conjuntos de dados distintos;
- Propor uma nova solução baseada nas técnicas do estado da arte e nos resultados das investigações realizadas;
- Comparar o método proposto com outras abordagens disponíveis;

- Documentar e publicar os resultados obtidos durante o desenvolvimento deste trabalho.

Como contribuição, este projeto visa ajudar na tomada de decisões em trabalhos de classificação de imagens, fornecendo *insights* sobre quais abordagens de aumentação de dados (individuais ou combinadas) melhor se adequam aos casos de uso abordados, sobre seu potencial para aplicação em outros cenários não tratados e provendo análises detalhadas sobre as vantagens e limitações de cada método investigado. Além disso, visamos propor uma nova metodologia de aumentação de dados que contribua para o desenvolvimento dessa área de pesquisas.

1.3 Questões de Pesquisa

Nesta seção, apresentaremos as questões de pesquisa que guiarão o desenvolvimento do nosso trabalho:

- É possível obter resultados comparáveis aos encontrados em trabalhos recentes de classificação de imagens, aumentando-se bases de dados que, originalmente, possuem uma pequena quantidade de amostras de treinamento?
- Pode-se realizar a combinação de diferentes técnicas de aumentação de dados, que sejam eficazes na superação de limitações presentes em cada abordagem individual, conduzindo à construção de melhores classificadores?
- É possível desenvolver uma metodologia de aumentação de dados com grande capacidade de generalização, de maneira que possa ser utilizada para melhorar o desempenho de classificadores treinados em bases de dados pertencentes a diferentes domínios de aplicação?
- Como as técnicas de aumentação de dados se comparam a outras abordagens que tratam o problema de escassez de amostras, tais como a regularização *dropout*, o pré-treinamento e novos paradigmas de aprendizado (como o *few-shot learning*)?

1.4 Organização do Texto

Este projeto está organizado em 6 capítulos. No Capítulo 1, apresentamos o problema de pesquisa a ser abordado durante o desenvolvimento deste trabalho, os principais objetivos e contribuições esperadas e as questões de pesquisa que guiarão este trabalho. No Capítulo 2, descrevemos os principais conceitos necessários para o entendimento deste projeto e apresentamos trabalhos relacionados relevantes encontrados na literatura. No Capítulo 3, apresentamos as bases de dados e os recursos computacionais a serem utilizados. No Capítulo 4, discutimos a metodologia e as métricas de avaliação que serão empregadas para o desenvolvimento e aperfeiçoamento dos nossos modelos. No Capítulo 5, apresentamos alguns resultados preliminares alcançados. Por fim, no Capítulo 6, apresentamos nosso plano de trabalho e cronograma.

Capítulo 2

Revisão Bibliográfica

Neste capítulo, apresentamos uma breve descrição de técnicas de aumentação de dados relevantes existentes na literatura, assim como trabalhos relacionados que as utilizam, principalmente para problemas de classificação de imagens quando em presença de pequenos conjuntos de treinamento. As técnicas aqui apresentadas são categorizadas em dois grupos: aquelas baseadas em métodos básicos e as baseadas em aprendizado profundo.

2.1 Métodos Básicos de Aumentação de Dados

Os métodos básicos de aumentação de dados são caracterizados por não serem baseados em abordagens modernas de aprendizado profundo.

As técnicas tradicionais podem ser categorizadas basicamente em dois tipos, as transformações geométricas e as fotométricas, sendo algumas de suas principais características a facilidade de implementação e o baixo custo computacional. Essas abordagens visam tratar, principalmente, os problemas decorrentes da existência de vieses de posição, iluminação e cor nos conjuntos de dados, que diminuem o poder de generalização dos modelos de aprendizado. Como exemplos de transformações geométricas, tem-se a rotação, a translação, a escala, o recorte e o espelhamento. Por sua vez, exemplos de transformações fotométricas são o *color jittering*, o ajuste de brilho e a equalização de histograma [72].

Existem outras abordagens que visam tratar, além dos vieses mencionados previamente, outros tipos, como os decorrentes da presença de ruídos e de oclusão de objetos da imagem. Como exemplos, tem-se a injeção de ruídos, que é caracterizada pela adição de valores aleatórios aos *pixels* de uma imagem, em que esses valores geralmente são obtidos a partir de uma determinada distribuição de probabilidades. Uma das vantagens dessa técnica é o incentivo para que os modelos aprendam características mais robustas. Outras técnicas incluem o emprego de filtros que alteram determinadas características da imagem, a aplicação de transformações elásticas, a utilização de operações de combinação de imagens e o uso de métodos que removem regiões do espaço de entrada (*pixels*) [72].

Perez et al. [61] realizaram a análise de treze abordagens de aumentação de dados para um problema de classificação de melanoma, exemplificadas na Figura 2.1. Estas incluem tanto métodos de aumentação tradicionais, tais como espelhamento, recorte aleatório, escala, e

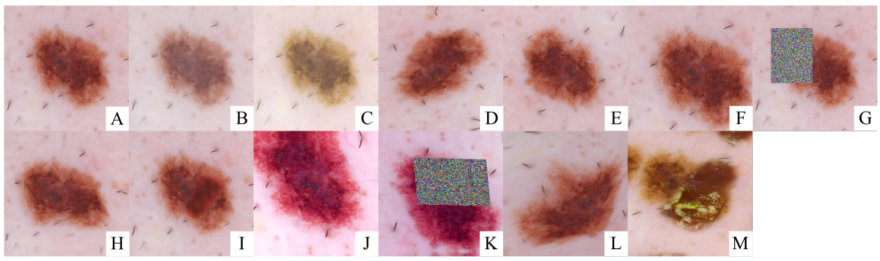


Figura 2.1: Exemplos dos métodos de aumentação de dados utilizados por Perez et al. [61].

mudanças de saturação, contraste, brilho e matiz; como também métodos menos usuais, tais como transformações elásticas, combinação de imagens e *Random Erasing* [95]. Além disso, algumas das abordagens são caracterizadas pela combinação de subconjuntos das outras.

Para os experimentos, foi utilizado o conjunto de dados fornecido pelo *ISIC Challenge 2017* [10], consistindo em 2000 imagens de treinamento, 150 imagens de validação e 600 imagens de teste. As técnicas de aumentação de dados foram aplicadas aos três conjuntos, além de também avaliar o conjunto de teste sem aumentações e a partir de 144 recortes de cada imagem [61]. Nos conjuntos de validação e teste, as aumentações e recortes foram utilizados para que a predição final sobre uma amostra original fosse obtida pela média das predições realizadas sobre as amostras a que ela deu origem. Além disso, três arquiteturas de redes neurais do estado da arte foram empregadas (Inception-v4 [78], ResNet-152 [29] e DenseNet-161 [33]), todas pré-treinadas na base de dados ImageNet [13], avaliando-se também o desempenho das aumentações para subconjuntos de diferentes tamanhos da base de dados original, contendo 125, 250, 500, 1000 e 1500 imagens.

Como resultados, observou-se que a aumentação dos dados de treinamento permite a construção de melhores modelos de classificação de melanoma, apesar de que algumas das operações adotadas podem piorar os resultados. Além disso, notou-se que o cálculo das médias das predições, no conjunto de teste, para as amostras geradas pelas abordagens de aumentação propostas e pela técnica de obtenção de 144 recortes sempre geram melhores resultados do que quando apenas as imagens de teste originais são utilizadas. Também, notou-se que a maior AUC (*Area Under the Curve*) foi alcançada ao empregar-se a abordagem de aumentação chamada de Conjunto Básico, tanto aos dados de treinamento como aos de teste (para cálculo das médias das predições), sendo composta por uma sequência de métodos tradicionais: recorte aleatório, cisalhamento, escala, espelhamento e alterações de saturação, contraste, brilho e matiz. O valor obtido foi de 0,882, superando o melhor resultado alcançado na competição (0,874).

Já com relação aos diferentes tamanhos de conjuntos de treinamento, observou-se que os modelos treinados com 500 ou mais amostras tiveram seus desempenhos bastante elevados ao serem empregadas a aumentação dos dados de treinamento e a média das predições sobre os dados aumentados do conjunto de teste. Por outro lado, a aplicação apenas da aumentação dos dados de treinamento gerou resultados piores do que aqueles obtidos sem aumentação, quando menos de 500 amostras foram utilizadas, e gerou apenas pequenas melhorias para os outros tamanhos dos conjuntos de treinamento.

Outro estudo comparativo foi realizado por Safdar et al. [71] em que oito métodos de

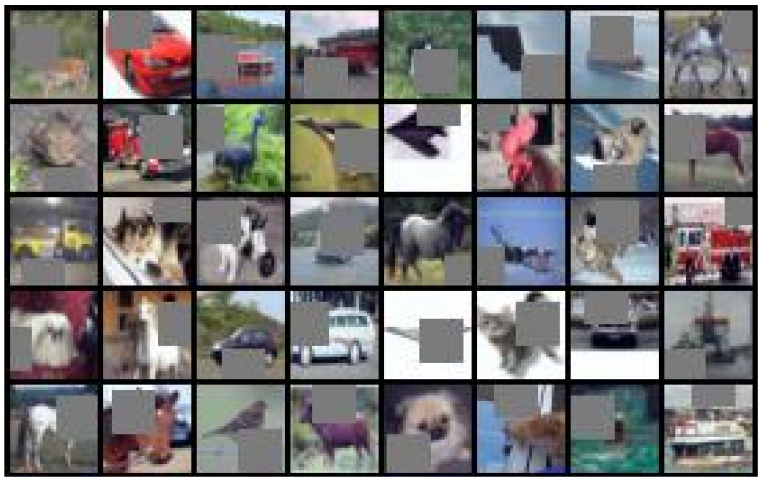


Figura 2.2: Exemplos do método de Cutout utilizado por DeVries et al. [15].

aumentação de dados básicos foram utilizados (espelhamento horizontal e vertical, adição de ruído, rotação de 90° e 180°, cisalhamento, recorte com escala e borramento) para a tarefa de classificação de tumores no cérebro, a partir de imagens de ressonância magnética, com o objetivo de identificar quais as abordagens mais adequadas para esse tipo de problema. Como resultado, foi observado que as técnicas de rotação levaram à construção dos modelos com as maiores acurácias.

DeVries et al. [15] desenvolveram um método, chamado de *Cutout*, que tem como objetivo incentivar a rede neural a aprender melhor o contexto completo da imagem, ao invés de utilizar apenas um subconjunto de características específicas que podem não estar presentes sempre, como nos casos em que há a oclusão de objetos da imagem. Ele consiste na remoção de uma região quadrada aleatória da imagem, conforme pode-se observar na Figura 2.2.

Essa técnica foi avaliada nas bases de dados CIFAR-10 [41], CIFAR-100 [41], SVHN [57] e STL-10 [9], alcançando-se o estado da arte nas três primeiras, com taxas de erro de 2,56%, 15,20% e 1,30%, respectivamente. Além disso, foram empregadas redes neurais ResNet-18 [29], Wide-ResNet [91] com profundidades de 28 e 16, e ResNets baseadas na regularização Shake-Shake [15]. Verificou-se também que a técnica *Cutout* pode ser empregada juntamente com outros métodos de regularização e de aumentação de dados.

Lopes et al. [50] desenvolveram uma técnica de aumentação de dados chamada de *Patch Gaussian*, que tem por objetivo melhorar a robustez e o poder de generalização dos modelos de aprendizado. Ela realiza a combinação de duas outras técnicas que, quando aplicadas individualmente, sofrem com um balanço entre as características que o método proposto visa melhorar: a injeção de Ruído Aditivo Gaussiano, que aumenta a robustez dos modelos ao ruído gaussiano, ao passo que pode diminuir a acurácia quando testados em dados limpos; e a técnica *Cutout* [15], que aumenta a acurácia dos modelos quando testados em dados limpos, mas não leva a um aumento de sua robustez.

Na técnica *Patch Gaussian*, uma região quadrada, de tamanho W×W, preenchida com um ruído gaussiano de desvio padrão σ, é adicionada a uma posição aleatória da imagem, de maneira que seu centro é sempre pertencente a esta. Na Figura 2.3, pode-se observar um exemplo de aplicação dessa técnica para um valor fixo de σ e diferentes valores de W.

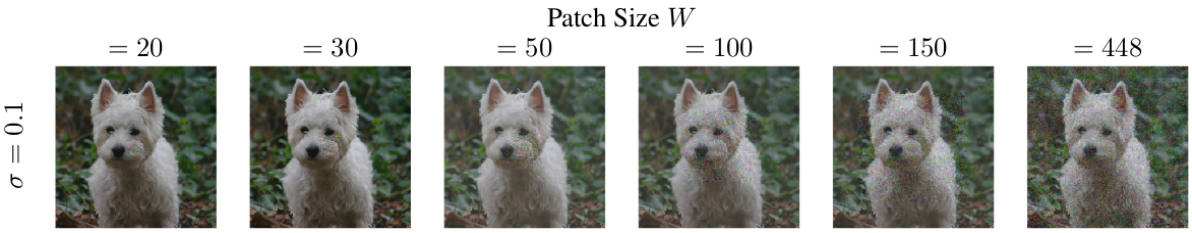


Figura 2.3: Exemplos do método de *Patch Gaussian* utilizado por Lopes et al. [50].

Para a validação da metodologia proposta, testes foram realizados utilizando-se os conjuntos de dados CIFAR-10 [41] e ImageNet [13], além das bases de dados CIFAR-C e ImageNet-C [30], que possuem imagens com corrupções encontradas no mundo real (causadas por ruídos, variações de brilho e diferentes condições climáticas), atingindo o estado da arte nas duas últimas. Foi mostrado que a metodologia desenvolvida atinge os objetivos de melhorar a robustez e a acurácia de teste dos modelos de aprendizado, reduzindo-se a sensibilidade dos modelos a ruídos de alta frequência e mantendo-se o aproveitamento de informações úteis também presentes em altas frequências. Além disso, demonstrou-se que essa técnica pode ser utilizada em conjunto com outros métodos de regularização e aumentação de dados, produzindo melhores resultados inclusive quando as imagens sofrem com outras perturbações diferentes de ruídos.

Summers et al. [76] desenvolveram uma técnica de combinação de imagens, chamada *Mixed-Example Data Augmentation*, que consiste em um conjunto de catorze operações não lineares que têm como objetivo gerar uma nova imagem a partir da combinação de duas outras imagens do conjunto de dados, conforme exemplificado na Figura 2.4.

Figura 2.4: Exemplos do método de *Mixed-Example* utilizado por Summers et al. [76].

Os experimentos foram realizados nas bases de dados CIFAR-10 [41], CIFAR-100 [41] e Caltech-256 [26]. Como resultados, foi observado que a maioria das operações propostas alcançou desempenho melhor do que os encontrados no estado da arte sem esse tipo de aumentação e do que aqueles obtidos pela combinação de imagens a partir de operações lineares, sendo as melhores operações aquelas que combinaram métodos não lineares e lineares. Como uma limitação, notou-se a necessidade de se aplicar métodos de aumentação tradicionais

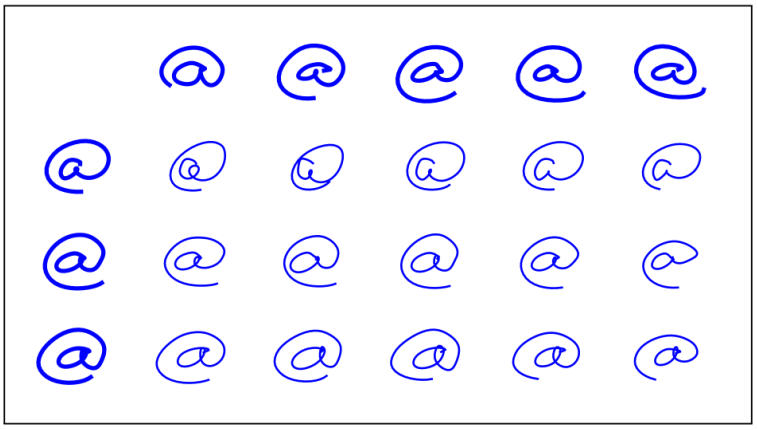


Figura 2.5: Exemplo de extrapolação do espaço de características usado por DeVries et al. [14].

antes do emprego da técnica proposta.

2.2 Aumentação Baseada em Aprendizado Profundo

Nesta seção, descreveremos alguns dos trabalhos relacionados que se baseiam em técnicas de aprendizado profundo para a realização da aumentação de dados. Serão tratados métodos que abordam a aumentação no espaço de características, que utilizam redes generativas adversariais, que adotam técnicas de transferência de estilo neural e que empregam técnicas de meta aprendizado.

2.2.1 Aumentação de Dados no Espaço de Características

O espaço de características profundo é caracterizado por ser uma representação de menor dimensionalidade, localizado nas camadas mais altas de uma rede neural profunda [72]. Certas direções nesse espaço correspondem a importantes transformações semânticas que podem ser exploradas durante o processo de aumentação dados.

DeVries et al. [14] desenvolveram uma técnica de aumentação de dados baseada no espaço de características e independente de domínio. Para isso, um Autoencoder LSTM (*Long Short-Term Memory*) [46] foi empregado, com duas camadas para o *encoder* e duas para o *decoder*, sendo treinado para produzir um vetor de características, chamado de vetor de contexto. Três tipos de aumentação de dados foram aplicadas, independentemente, a esse vetor: a adição de ruído gaussiano aleatório, a interpolação e a extrapolação entre pares de vetores. Na Figura 2.5, pode-se observar um exemplo de aumentação empregando a técnica de extrapolação dos vetores de contexto, em que as imagens em negrito representam as entradas e as restantes são as amostras produzidas.

Os experimentos foram realizados tanto em conjuntos de dados com amostras representando séries temporais, tais como fala, sensores e captura de movimento, assim como em bases de dados com amostras estáticas (MNIST [43] e CIFAR-10 [41]). Ainda na base de dados CIFAR-10, duas abordagens de teste foram experimentadas: o teste utilizando os vetores

de contexto produzidos pelo autoencoder e sujeitos às aumentações de dados propostas; e o teste empregando as imagens reconstruídas a partir desses vetores.

Como resultados, foi observado que a extrapolação dos vetores de contexto produzidos pelo autoencoder obteve o melhor desempenho, provavelmente por aumentar a variabilidade do conjunto de dados. Apesar disso, em cenários em que as fronteiras de decisão são menos complexas (como em amostras linearmente distribuídas), a interpolação pode ser uma melhor alternativa. Além disso, verificou-se que o desempenho dos modelos piora ao serem treinados com as imagens reconstruídas a partir dos vetores de contexto. Mesmo assim, foi mostrado que a metodologia desenvolvida, em conjunto com arquiteturas de redes mais sofisticadas, pode ser um complemento às técnicas de aumentação de dados usuais.

Wang et al. [84] desenvolveram um método de aumentação de dados no espaço de características, chamado *Implicit Semantic Data Augmentation* (ISDA), que tem como base o fato de que, no espaço de características profundo, certas direções correspondem a transformações semânticas importantes. Nele, durante o treinamento do modelo, uma matriz de covariância é calculada para cada classe do conjunto de dados, sendo utilizada para a captura das variações e, consequentemente, das direções semânticas mais importantes dentro de cada classe. Essas matrizes também são empregadas para a obtenção de vetores aleatórios, amostrados a partir de uma distribuição normal, que podem ser usados para a aumentação de dados. Além dessa técnica de obtenção de transformações, uma das características importantes dessa metodologia é o desenvolvimento de uma função de perda mais robusta, que permite o treinamento de modelos sem a necessidade de se criar, explicitamente, novas amostras.

Para a validação do método desenvolvido, três bases de dados foram usadas (CIFAR-10, CIFAR-100 [41] e ImageNet [13]), além de diferentes arquiteturas de redes neurais (ResNet [29], SE-ResNet [31], Wide-ResNet [91], ResNeXt [89] e DenseNet [32]), métodos de aumentação do estado da arte (Cutout [15] e AutoAugment [12]) e *baselines* (*Dropout* [75], *Large-Margin Softmax Loss* [49], *Disturb Label* [88], *Focal Loss* [47], *Center Loss* [86], L_q *Loss* [93] e métodos de aumentação semântica baseados em GANs).

Como resultados, foi observado que a ISDA melhora o desempenho de todos os modelos de aprendizado empregados, especialmente quando a base de dados possui menos amostras por classe. Além disso, a combinação da ISDA com outros métodos de aumentação não semântica, tais como Cutout e AutoAugment, leva a uma melhora ainda maior da eficácia dos modelos, diminuindo os erros de classificação do conjunto de teste. Também, observou-se que a ISDA apresenta resultados competitivos com todos os *baselines* comparados e que as amostras que podem ser geradas a partir dessa metodologia possuem informações semânticas coerentes, sendo esta capaz de transformar características interessantes, como o plano de fundo, os ângulos de visão, as cores e os tipos dos objetos nas imagens, o que não é possível com os métodos de aumentação tradicionais. Na Figura 2.6, pode-se observar exemplos de transformações que podem ser realizadas com essa técnica.

2.2.2 Aumentação Baseada em Redes Adversariais Generativas

Redes Adversariais Generativas (do inglês, *Generative Adversarial Networks* - GANs) constituem um tipo de modelo generativo que vêm alcançando grande destaque e mostram-se extre-

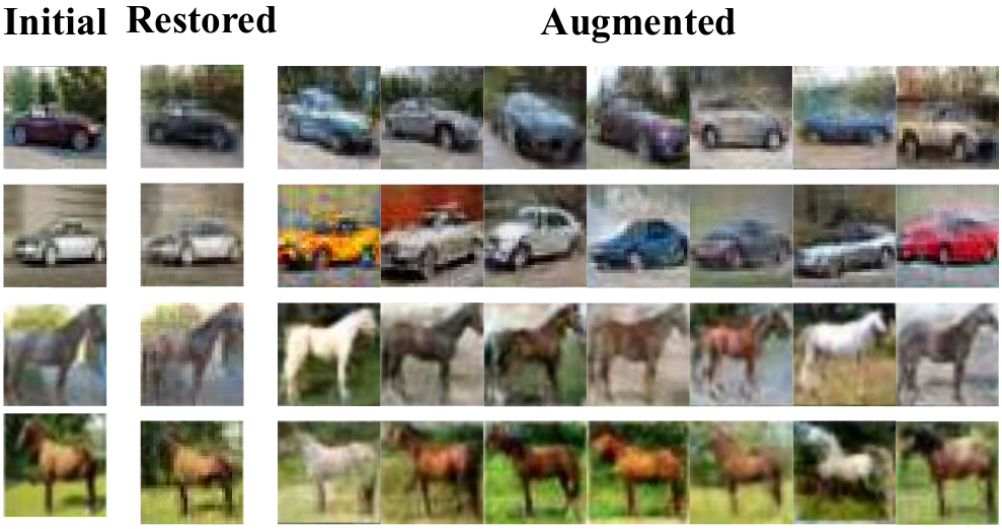


Figura 2.6: Exemplos do método empregado por Wang et al. [84].

mamente promissoras no contexto de aumentação de dados [72]. Embora diversas variações desse modelo venham sendo desenvolvidas, de maneira a superar algumas de suas limitações e melhorar seu desempenho em determinados domínios, uma GAN pode ser vista como um modelo composto por duas redes neurais: uma chamada de geradora, que é responsável por produzir as amostras sintéticas e tem como principal objetivo enganar a discriminadora; esta, por sua vez, tem como tarefa identificar se as amostras que recebe como entrada provêm da distribuição de dados original ou são sintéticas.

Durante o treinamento de uma GAN, tanto a geradora como a discriminadora melhoram em suas tarefas, de maneira que, ao final desse processo, a geradora se torna capaz de produzir amostras que enganam a discriminadora com um elevado grau de confiança. Para isso, a GAN necessita aprender a distribuição de dados original [72].

Frid-Adar et al. [21] utilizaram duas técnicas de aumentação de dados, baseadas em GANs, para o problema de classificação de lesões no fígado: a *Deep Convolutional* GAN (DCGAN) [63], que tem como característica o emprego de camadas convolucionais tanto para a criação da rede geradora como da discriminadora; e a *Auxiliary Classifier* GAN (ACGAN) [60], que é semelhante à DCGAN, mas utiliza os rótulos das amostras de treinamento como uma entrada condicional do modelo de aumentação, e a discriminadora, além de diferenciar dados reais daqueles sintetizados, também realiza a predição de suas classes. No trabalho desenvolvido, foi construída uma DCGAN para cada classe do conjunto de dados e uma única ACGAN capaz de gerar amostras sintéticas para qualquer uma das classes.

Para a avaliação dos métodos desenvolvidos, utilizou-se uma base de dados de tomografias computadorizadas (CT) de lesões no fígado, que contém 53 imagens de cistos, 64 imagens de metástases e 65 imagens de hemangiomas. Além disso, uma CNN desenvolvida pelos autores foi empregada para a tarefa de classificação, e realizou-se experimentos tanto com conjuntos de dados resultantes apenas de aumentações tradicionais, como também com amostras produzidas pela combinação de métodos tradicionais com as técnicas de geração de dados sintéticos desenvolvidas.

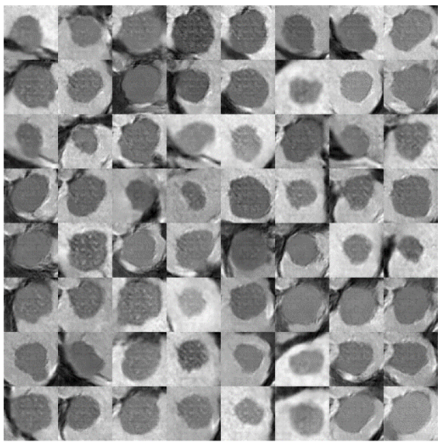


Figura 2.7: Exemplos de amostras de cistos produzidas pela rede DCGAN [21].

Como resultados, observou-se que o emprego dos dados gerados pelos métodos desenvolvidos, em conjunto com os obtidos a partir de aumentações tradicionais, permitiu a construção de modelos de maior desempenho, com a DCGAN produzindo melhores resultados do que a ACGAN. Também, foi observado que as amostras geradas possuem tanto aparência como também características significativas, mostrando a eficiência dos métodos apresentados e alcançando resultados do estado da arte. Na Figura 2.7, é possível observar exemplos de amostras da classe Cisto geradas pelo modelo DCGAN utilizado por Frid-Adar et al. [21].

Zhang et al. [92] desenvolveram uma metodologia de aumentação baseada em GANs, chamada de *Deep Adversarial Data Augmentation* (DADA), para problemas de classificação com regime extremamente baixo de dados. Ela é composta de duas partes: uma GAN condicionada à classe, que é treinada de maneira supervisionada; e uma função de perda, chamada de *2k loss*, que tem como objetivo fortalecer a influência tanto das amostras reais, como das amostras sintéticas, no estabelecimento das fronteiras de decisão.

Para os experimentos, foram empregadas as bases de dados CIFAR-10 [41], restringindo-se a quantidade de amostras utilizadas, e a base *Curated Breast Imaging Subset of the Digital Database for Screening Mammography* (CBIS-DDSM) [44], representando um conjunto pequeno de dados do mundo real, utilizado para a classificação de tumores. Testes foram realizados empregando-se o conjunto de dados sem aumentação, com aumentação tradicional, apenas com a DADA e com a combinação entre os métodos tradicionais e a DADA.

Como resultados, na base de dados CIFAR-10, em que apenas 10% das amostras originais foram utilizadas, a combinação de métodos tradicionais de aumentação com a metodologia desenvolvida obteve a melhor eficácia. Com relação à base de dados CBIS-DDSM, tanto a DADA sozinha, como a DADA com aumentação tradicional, obtiveram resultados do estado da arte, sendo a segunda a de maior eficácia (65,49% de acurácia).

Zhu et al. [98] desenvolveram um *framework* para a tarefa de classificação de sentimentos a partir de imagens, o qual consiste em uma CNN para a classificação das imagens e uma *Cycle-Consistent Adversarial Network* (CycleGAN) [97] para a aumentação dos conjuntos de dados pela geração de amostras sintéticas. Sete classes foram consideradas, sendo suas distribuições desbalanceadas: neutralidade, medo, raiva, aversão, tristeza, felicidade e surpresa. Além

Figura 2.8: Exemplos de amostras geradas por Zhu et al. [98].

disso, a CycleGAN foi adaptada para utilizar a função de perda *least-squared loss* [54] em combinação com a *adversarial loss* original, de maneira a tratar o problema de *vanishing gradients* [3]. A capacidade dessa rede adversarial de realizar o mapeamento entre domínios foi aproveitada para a criação de amostras sintéticas para as categorias minoritárias (alvo) a partir das imagens existentes nas classes majoritárias (referência).

Para a validação do método proposto, três bases de dados foram empregadas: a *Facial Expression Recognition Database* (FER2013) [25], a *Static Facial Expressions in the Wild* (SFEW) [16] e a *Japanese Female Facial Expression* (JAFFE) [51]. Experimentos foram realizados empregando-se a CNN proposta na criação de uma *baseline*, em que o treinamento foi realizado sem a aumentação de dados. Também, criou-se modelos em que apenas duas das classes (tristeza e aversão) foram aumentadas. Por fim, todas as classes foram aumentadas, exceto a de neutralidade, pois esta foi utilizada como a referência. É importante destacar que os modelos treinados nos conjuntos SFEW e JAFFE foram pré-treinados na base FER2013, devido à sua pequena quantidade de amostras.

Como resultados, observou-se que as aumentações empregadas melhoraram os desempenhos dos modelos em relação ao *baseline*, com as bases FER2013, SFEW e JAFFE apresentando acurácias iguais a 94,71%, 39,07% e 95,80%, respectivamente. Além disso, as acurácias para todas as classes (não somente para aquelas que foram aumentadas) também melhoraram. Ademais, foi observado que a técnica de aumentação desenvolvida melhorou a distribuição dos dados e tornou mais claras as fronteiras de decisão entre as classes. Na Figura 2.8, pode-se observar exemplos de amostras geradas pela CycleGAN para cada uma das classes, exceto para neutralidade (as imagens mais à esquerda são as originais).

2.2.3 Aumentação Baseada em Transferência de Estilo Neural

A Transferência de Estilo Neural consiste na construção de um algoritmo capaz de transferir o estilo de uma imagem para outra imagem, mantendo o conteúdo semântico da última. Os

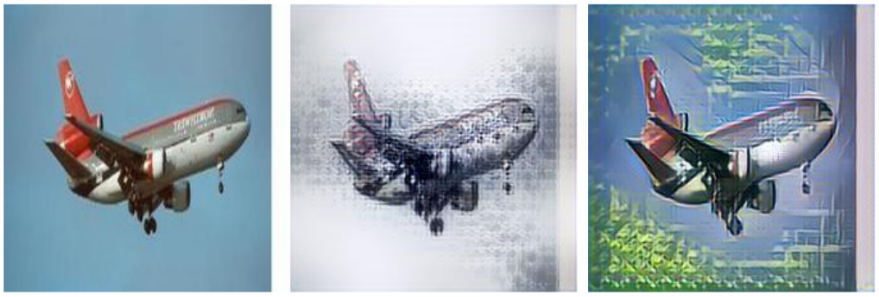


Figura 2.9: Exemplos de transferências de estilo aplicadas por Zheng et al. [94].

algoritmos do estado da arte desta técnica podem ser divididos em dois grupos: os descritivos, que realizam a transformação iterativa de uma imagem de ruído, de maneira que, ao final desse processo, ela contenha o conteúdo de uma imagem e o estilo de outra; e os generativos, que atingem o mesmo resultado dos anteriores, mas, ao invés de realizarem transformações iterativas, empregam um modelo pré-treinado para a transformação de estilo desejada, tornando mais eficiente esse processo [94]. Quando aplicada à tarefa de aumentação de dados, a transferência de estilo neural é similar a transformações de cor, permitindo também o emprego de texturas e estilos artísticos diferentes [72].

Zheng et al. [94] investigaram a efetividade da transferência de estilo neural como uma abordagem de aumentação de dados para tarefas de classificação de imagens. Para isso, oito imagens com diferentes estilos foram escolhidas como referência; as transformações foram realizadas utilizando-se a arquitetura proposta por Engstrom [18], consistindo em uma rede responsável por aprender o estilo e outra encarregada de calcular as funções de perda que atualizam a primeira; e as arquiteturas VGG16 e VGG19 [73] foram adotadas para a tarefa de classificação das amostras.

Para o treinamento da rede de transferência de estilo, empregou-se a base dados COCO 2014 [48], e um modelo de aumentação foi criado para cada um dos oito estilos escolhidos (*Snow*, *RainPrincess*, *Scream*, *Wave*, *Sunflower*, LAMuse, Udnie e *Your Name*). Já para a tarefa de classificação, as bases Caltech-101 [20] e Caltech-256 [26] foram utilizadas para a construção dos modelos. Além disso, para a análise da efetividade da metodologia investigada, testes foram realizados apenas com os dados originais (*baseline*), apenas com a aplicação de aumentações de dados tradicionais, apenas com as aumentações resultantes da transferência de estilo e com a combinação das duas últimas. Na Figura 2.9, pode-se observar um exemplo de aplicação do método proposto, em que a imagem à esquerda é a original, a do meio é o resultado da transferência do estilo *Snow* e a da direita é resultante do estilo *Your Name*.

Como resultados, observou-se que a transferência de estilo neural é uma abordagem efetiva de aumentação de dados para a melhoria de modelos de classificação de imagens, obtendo uma elevação de acurácia, em relação ao *baseline*, de 1,92% para o modelo VGG16 e de 1,31% para o VGG19. Também, foi observado que a combinação dessa metodologia com técnicas tradicionais de aumentação pode melhorar ainda mais o desempenho dos modelos. Entretanto, uma limitação ainda encontrada na aumentação utilizando a transferência de estilo neural é a escolha de quais estilos utilizar.

Figura 2.10: Exemplos de transferências de estilo aplicadas por Jackson et al. [36].

Jackson et al. [36] desenvolveram uma metodologia de aumentação de dados, chamada *Style Augment*, que se baseia na transferência de estilos neurais aleatórios. Nela, uma rede de transferência capaz de aplicar uma grande variedade de estilos, mesmo de domínios não vistos durante o treinamento (desenvolvida por Ghiasi et al. [24]), é adaptada para que seja capaz de utilizar estilos provenientes de vetores de características aleatórios.

Para isso, vetores de características de estilo (*style embeddings*) são amostrados a partir de uma distribuição normal, em que a média e a matriz de covariância utilizadas para sua construção são obtidas a partir da base de dados *Painter By Numbers* (PBN) [58], também utilizada para o treinamento da rede de transferência. O emprego de *style embeddings* simula a escolha aleatória de uma imagem a partir dessa base de dados, mas torna mais eficiente o processo de obtenção de estilo. Na Figura 2.10, é possível observar um exemplo de aplicação de estilos aleatórios a uma amostra, em que a imagem na parte superior esquerda é a original e as demais são resultantes das aumentações.

Para a validação da metodologia proposta, três tarefas de visão computacional foram escolhidas: a classificação de imagens, em que se empregou o conjunto de dados STL-10 [9] para o treinamento e teste de um classificador Inception-v3 [79]; a classificação de imagens pertencentes a diferentes domínios, em que se utilizou a base dados *Office dataset* [70] e os modelos Inception-v3 [79], ResNet-18 [29], ResNet-50 [29] e VGG16 [73]; e a estimativa de profundidade monocular, em que se utilizou uma rede U-Net [4], com 65000 imagens sintéticas obtidas a partir de ambientes virtuais de jogos [69] para treinamento, e o conjunto de teste da base KITTI [82] para a avaliação. Além disso, quatro abordagens de experimentação foram adotadas: sem aumentação de dados, com aumentações tradicionais, com aumentações de estilo e com a combinação das duas últimas.

Como resultados, para a tarefa de classificação de imagens, observou-se que a metodologia proposta melhorou a acurácia do modelo com relação aos resultados sem aumentação. Na classificação de imagens pertencentes a domínios diferentes, em alguns casos a transferência de estilo aleatório superou a acurácia das técnicas tradicionais, mostrando sua efetividade e capacidade de inserir invariâncias que amenizam o *overfitting* dos modelos. Já com relação à estimativa de profundidade monocular, os modelos treinados com a abordagem proposta apresentaram melhores desempenhos (com relação a métricas de erro e de acurácia) do que os obtidos com técnicas de aumentação tradicionais. Para todos os três tipos de tarefa, a combinação de métodos tradicionais com a abordagem proposta gerou os melhores resultados, superando o estado da arte na classificação de imagens da base STL-10 (80,8%) quando amostras não rotuladas não são empregadas.

2.2.4 Aumentação de Dados Baseada em Meta-Aprendizado

O meta-aprendizado tem como objetivo a aprendizagem de dois modelos, um deles responsável por realizar uma determinada tarefa (como a classificação de imagens) e o outro responsável pela otimização do primeiro e pela sua atualização quando dados de uma nova tarefa são recebidos [66]. Diversos trabalhos de meta-aprendizado vêm sendo desenvolvidos, tanto para o projeto de arquiteturas que levam às melhores acurácias em determinadas tarefas [99], como para o tratamento de problemas em que poucas amostras estão disponíveis [77]. Também, diversas abordagens de otimização vem sendo experimentadas, tais como algoritmos evolucionários e aprendizado por reforço. Apesar disso, essas abordagens ainda enfrentam algumas adversidades, como a dificuldade de implementação dos modelos.

Cubuk et al. [12] desenvolveram um método de aumentação de dados baseado em meta-aprendizado, chamado AutoAugment, que tem como objetivo a automatização da escolha das estratégias de aumentação. Para isso, um algoritmo de busca baseado em aprendizado por reforço e um espaço de busca discreto são empregados.

O espaço de busca é composto por um conjunto de dezesseis operações de transformação de imagens (Cisalhamento X/Y, Translação X/Y, Rotação, AutoContraste, Inversão, Equalização, Solarização, Posterização, Contraste, Cor, Brilho, Nitidez, Cutout [15] e SamplePairing [34]) que, juntamente com parâmetros de probabilidade e magnitude dessas operações, são combinadas para formar sub-políticas que, por sua vez, são unidas para formar políticas (estratégias de aumentação). Para se determinar quais são as melhores políticas de aumentação para um determinado conjunto de dados, uma rede neural recorrente (do inglês, *Recurrent Neural Network* - RNN) [100], chamada de controlador, é treinada para escolher as operações que levam um modelo de classificação auxiliar (*child*) a obter a melhor acurácia de validação. Em cada iteração de treinamento, a acurácia é enviada para o controlador na forma de uma recompensa, que determina a atualização do classificador.

Para a validação do método proposto, duas abordagens foram experimentadas: o cálculo e aplicação de políticas de aumentação específicas para o conjunto de dados de destino (*AutoAugment-direct*) e o uso de políticas, aprendidas em um determinado conjunto de dados, em outro conjunto de destino (*AutoAugment-transfer*). No primeiro caso, foram empregadas as bases de dados CIFAR-10 [41], CIFAR-100 [41], SVHN [57] e ImageNet [13], com as arquiteturas Wide-ResNet-28-10 [91], Shake-Shake [22], ShakeDrop [90] e AmoebaNet [68]. Já no caso de transferência de políticas, utilizou-se as estratégias aprendidas na base de dados ImageNet para o treinamento e teste de modelos Inception-v4 [78] nas bases Oxford 102 Flowers [59], Caltech-101 [20], Oxford-IIIT Pets [17], FGVC Aircraft [52] e Stanford Cars [39]. Além disso, comparou-se a abordagem proposta com outro método de aumentação automatizada proposta por Ratner et al. [67].

Como resultados, observou-se que as políticas geradas pelo AutoAugment conduziram às maiores eficácias ao empregar-se a abordagem direta, com taxas de erro de 1,5%, 10,7% e 1,0% para as bases CIFAR-10, CIFAR-100 e SVHN, respectivamente, e acurácia *top* 1 de 83,5% para a base ImageNet, atingindo-se o estado da arte nesses conjuntos. Também, notou-se que o impacto positivo desse método é maior em conjuntos de dados com poucas amostras, e que os tipos de aumentação mais frequentes diferem de acordo com a base, sendo os da ImageNet e CIFAR-(10/100) mais relacionados à cor e os da SVHN mais ligados

a transformações geométricas. Com respeito à transferência de políticas, o AutoAugment também gerou os melhores valores, mostrando que esta abordagem pode encontrar estratégias mais genéricas que podem ser aplicadas a problemas diferentes, apesar de que as políticas aprendidas sobre distribuições próximas das de destino acarretam os melhores desempenhos. Por fim, o AutoAugment também apresentou melhores resultados do que aqueles obtidos pelo método de aumentação proposto por Ratner et al. [67].

Na Figura 2.11, é possível observar um exemplo de política de aumentação aplicada sobre uma determinada amostra da base ImageNet, em que apresenta-se o tipo de operação, sua probabilidade de ocorrência e magnitude, para cada sub-política que compõe essa política.

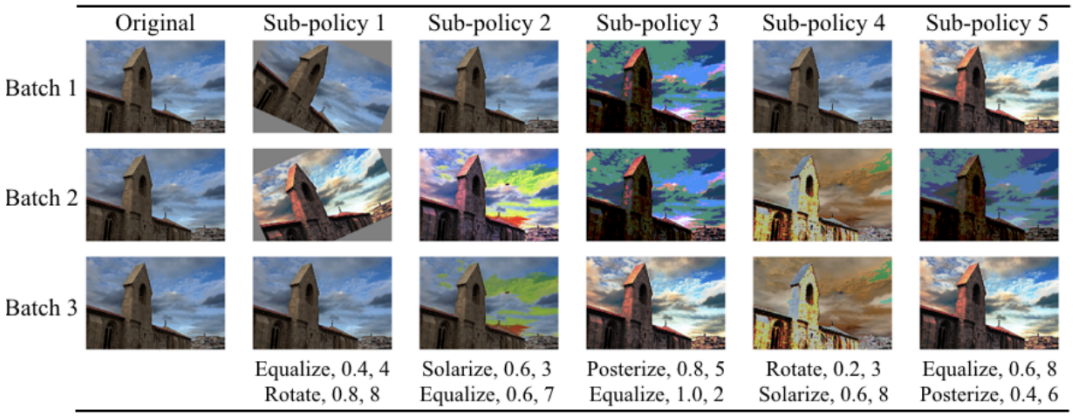


Figura 2.11: Exemplo de política de aumentação aplicada por Cubuk et al. [12].

Geng et al. [23] propuseram um método de aumentação de dados, chamado ARS-aug, que tem como objetivo a adaptação do método AutoAugment [12] pela substituição do algoritmo de aprendizado por reforço, empregado na busca das operações de aumentação, pelo algoritmo *Augmented Random Search* [53], de maneira a trocar o espaço de busca discreto por um espaço contínuo. Com isso, procurou-se melhorar o desempenho da busca, mantendo a diversidade das políticas.

Para validar a metodologia proposta, testes foram realizados nas bases de dados CIFAR-10 [41], CIFAR-100 [41] e ImageNet [13], atingindo-se o estado da arte, com um erro de 1,26% e 10,24% nos conjuntos CIFAR-10 e CIFAR 100, e uma acurácia *top* 1 de 83,88% na base ImageNet. Além disso, foi observado que a busca em um espaço contínuo permitiu a otimização das probabilidades, favorecendo a escolha das operações mais significativas para o problema tratado e a diversidade das políticas, e que a magnitude tornou-se mais precisa, fornecendo melhores indicações da influência de cada operação.

Mihn et al. [56] desenvolveram um *framework* para aumentação de dados automatizada, baseado em uma abordagem de aprendizado por reforço, cujo objetivo é o aprendizado de uma sequência de transformações ótima para cada imagem do conjunto de dados. Esse *framework* é composto de dois elementos: um agente, que possui uma rede neural profunda para a geração das possíveis ações (classificar imagem ou aplicar política de transformação) e um tomador de decisões que decide parar ou continuar a transformação de uma imagem; e o ambiente, que é responsável por aplicar a política escolhida pelo agente (sequência de transformações), e por retornar a imagem transformada e uma recompensa que indica o

impacto da transformação.

Para a avaliação da abordagem proposta, uma rede DDQN (*Dueling Deep Q-Network*) [85], com três diferentes arquiteturas propostas pelos autores, é utilizada para compor o agente; avalia-se a acurácia dessas redes quando treinadas apenas com os dados originais (*baseline*), quando treinadas com os dados originais através do *framework* de aprendizado por reforço proposto, e quando treinadas desta última maneira e refinadas nas imagens resultantes das transformações. Além disso, quatro bases de dados, com diferentes graus de ruído, foram empregadas: MNIST [43], SVHN [57], CIFAR-10 [41] e DOGCAT [1].

Como resultados, observou-se que o *framework* proposto permitiu a construção de modelos mais robustos e com maiores acurácias do que o *baseline*, além de fornecer um processo que facilita a análise e explicabilidade das políticas escolhidas. Apesar disso, resultados do estado da arte não foram alcançados, indicando a necessidade de escolha de melhores arquiteturas de redes neurais para compor o agente de aprendizado por reforço. Também, o emprego de outras técnicas de aumentação para compor o espaço de busca precisa ser investigado, uma vez que apenas dois métodos básicos (rotação e espelhamento) foram experimentados.

2.3 Resumo dos Trabalhos Relacionados

A Tabela 2.1 apresenta um resumo dos trabalhos correlatos, de acordo com as seguintes características:

- Invariâncias inseridas: indica alguns dos tipos de invariâncias que o método de aumentação proposto é capaz de adicionar ao conjunto de dados.
- Natureza do método: neste projeto, dividimos os métodos apresentados nos trabalhos relacionados em básicos ou baseados em aprendizado profundo.
- Tipo de aumentação: as técnicas de aumentação podem ser categorizadas em dois tipos básicos, que são os que inflam o conjunto de dados através de distorções aplicadas às imagens originais ou por meio da geração de amostras sintéticas.
- Espaço de aumentação: indica se as técnicas de aumentação são aplicadas ao espaço de entrada (imagem) ou ao espaço de características.
- Natureza das transformações: indica se um conjunto fixo de aumentações é sempre aplicado ou se esse conjunto pode variar.
- Escolha das transformações: indica se a escolha das transformações é feita de acordo com todas as amostras, com cada classe ou especificamente para cada amostra.
- Usa outros métodos: indica se, além do método proposto no trabalho, outras abordagens de aumentação são empregadas.

Conforme pode-se observar, cada método discutido possui suas vantagens e limitações, como o tipo de invariâncias que eles inserem no conjunto de dados. A maioria deles foi experimentada com outras abordagens de aumentação, mostrando que podem ser combinadas para

Tabela 2.1: Resumo dos trabalhos relacionados.

Trabalho	**Invariâncias inseridas**	**Natureza do método**	**Tipo de aumentação**	**Espaço de aumentação**	**Natureza das transformações**	**Escolha das transformações**	**Usa outros métodos**
Perez et al. [61]	Posição, Escala, Cor, Oclusão, Iluminação, Ângulo de Visão	Básico	Distorção e Super Amostragem	Espaço de Entrada	Transformações pré-determinadas	Para toda a base de dados	Sim
DeVries et al. [15]	Oclusão	Básico	Distorção	Espaço de Entrada	Transformações pré-determinadas	Para toda a base de dados	Sim
Lopes et al. [50]	Oclusão, Ruído e outras decorrentes de corrupções comuns [30]	Básico	Distorção	Espaço de Entrada	Transformações pré-determinadas	Para toda a base de dados	Sim
Summers et al. [76]	Decorrentes da combinação não linear de imagens	Básico	Super Amostragem	Espaço de Entrada	Transformações pré-determinadas	Para toda a base de dados	Sim
DeVries et al. [14]	Decorrente da interpolação, extrapolação e adição de ruídos ao espaço de características	Aprendizado Profundo	Super Amostragem	Espaço de Características	Transformações pré-determinadas	Para toda a base de dados	Sim
Wang et al. [84]	Plano de Fundo, Cor, Ângulo de Visão, Escala, Iluminação, Posição, Oclusão, Tipo de Objetos	Aprendizado Profundo	Super Amostragem	Espaço de Características	Transformações variáveis	Para cada classe	Sim
Frid-Adar et al. [21]	Decorrentes do emprego de métodos generativos	Aprendizado Profundo	Super Amostragem	Espaço de Entrada	Transformações variáveis	Para cada classe e para toda a base de dados	Sim
Zhang et al. [92]	Decorrentes do emprego de métodos generativos	Aprendizado Profundo	Super Amostragem	Espaço de Entrada	Transformações variáveis	Para toda a base de dados	Sim
Zhu et al. [98]	Decorrentes do emprego de métodos generativos	Aprendizado Profundo	Super Amostragem	Espaço de Entrada e Espaço de Características	Transformações variáveis	Para cada classe	Não
Zheng et al. [94]	Cor, Iluminação, Textura e Estilo	Aprendizado Profundo	Distorção	Espaço de Entrada	Transformações pré-determinadas	Para toda a base de dados	Sim
Jackson et al. [36]	Cor, Iluminação, Textura e Estilo	Aprendizado Profundo	Distorção	Espaço de Entrada e Espaço de Características	Transformações variáveis	Para toda a base de dados	Sim
Cubuk et al. [12]	Posição, Oclusão, Ângulo de Visão, Iluminação, Cor, Forma, Nitidez e as decorrentes da combinação de imagens	Aprendizado Profundo	Distorção e Super Amostragem	Espaço de Entrada	Transformações variáveis	Para toda a base de dados	Sim
Geng et al. [23]	Posição, Oclusão, Ângulo de Visão, Iluminação, Cor, Forma, Nitidez e as decorrentes da combinação de imagens	Aprendizado Profundo	Distorção e Super Amostragem	Espaço de Entrada	Transformações variáveis	Para toda a base de dados	Sim
Mihn et al. [56]	Ângulo de Visão	Aprendizado Profundo	Distorção	Espaço de Entrada	Transformações variáveis	Para cada amostra	Não

a obtenção de desempenhos ainda melhores. Também, apesar de grande parte dos métodos empregar as mesmas transformações em todas as amostras, alguns deles mostram o potencial

de se construir transformações específicas para cada classe ou para cada amostra. Além disso, observou-se que, embora aumentações baseadas em um conjunto pré-determinado de transformações possam trazer benefícios para um modelo, o emprego de técnicas que aprendem o melhor conjunto de transformações para a tarefa em questão pode gerar resultados ainda melhores.

Por fim, além da aumentação de dados, várias outras técnicas têm sido desenvolvidas para o tratamento do problema de *overfitting*. Dentre elas, estão a criação de novos paradigmas de aprendizado (como *few-shot learning* [77]), de métodos de regularização (por exemplo, *dropout* [75]), assim como técnicas de transferência de aprendizado e pré-treinamento. Gururangan et al. [28] investigaram o efeito do pré-treinamento, tanto adaptado para domínio como para tarefa, no desempenho de modelos de processamento de linguagem natural. Eles mostraram que o uso dessas abordagens traz melhorias, principalmente quando são combinadas.

Capítulo 3

Materiais

Neste capítulo, descrevemos as bases de dados e os recursos computacionais que serão utilizados no desenvolvimento deste projeto.

3.1 Bases de Dados

Este projeto utilizará, inicialmente, quatro bases de dados públicas, correspondentes a quatro domínios de aplicação diferentes: VGGFace2 [7] para reconhecimento de faces, Places365 - Standard [96] para reconhecimento de cenas, Digipathos Embrapa [5,6] para identificação de espécies doentes em plantações e COVIDx [83] para a identificação de pacientes contaminados com o novo coronavírus SARS-CoV-2 (COVID-19). A seguir, cada uma dessas bases de dados é descrita.

3.1.1 VGGFace2

A base VGGFace2 [7] possui, aproximadamente, 3,3 milhões de imagens de faces obtidas a partir de 9.131 celebridades e figuras públicas, abrangendo várias etnias, profissões, poses, idades, condições de iluminação e planos de fundo. As dimensões das imagens variam de valores menores do que 50 *pixels* (px) até valores maiores do que 300 px, de forma que os conjuntos de treinamento e teste possuem tamanhos próximos de 36 GB e 19 GB, respectivamente. Além disso, todas as imagens estão no formato JPEG (*Joint Photographic Experts Group*), a quantidade de imagens por classe varia de 87 a 843 e as amostras são divididas de maneira que o conjunto de treinamento contém 8.631 das classes e o conjunto de teste possui as 500 restantes.

Ademais, as amostras da base VGGFace2 [7] foram coletadas a partir da ferramenta de buscas de imagens do Google. Há imagens coloridas e monocromáticas. Em algumas delas, há indivíduos utilizando acessórios como óculos e chapéus, e há situações em que uma parte da face está ocluída por algum obstáculo. Na Figura 3.1, pode-se observar exemplos de classes presentes nesta base, mostrando indivíduos sob diferentes poses e idades.

(a) Exemplo 1: Anne Schedeen. (b) Exemplo 2: Anne Schedeen. (c) Exemplo 3: Anne Schedeen.

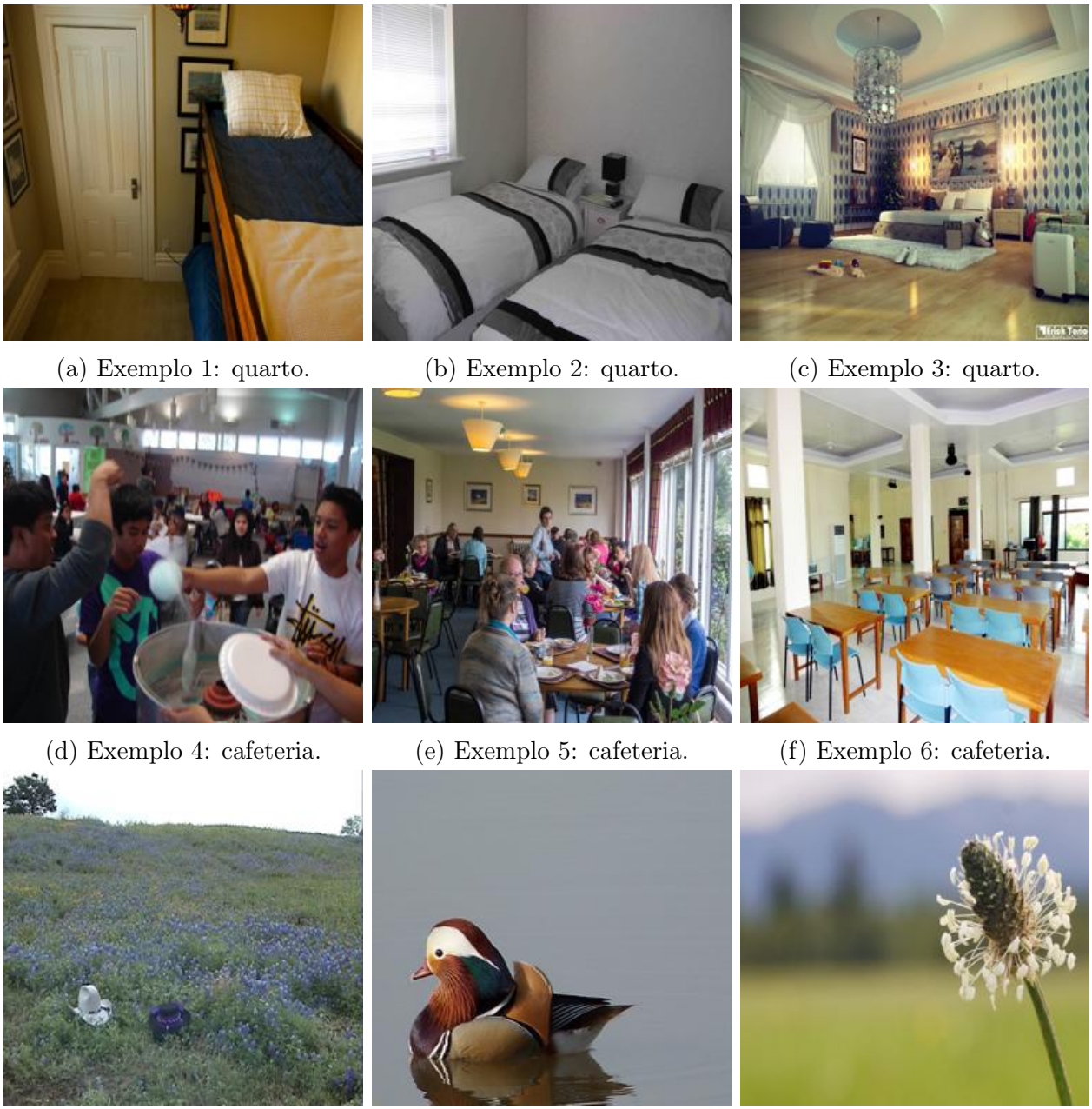
(d) Exemplo 4: Felipe Alou. (e) Exemplo 5: Felipe Alou. (f) Exemplo 6: Felipe Alou.

(g) Exemplo 7: Lang Ping. (h) Exemplo 8: Lang Ping. (i) Exemplo 9: Lang Ping.

Figura 3.1: Amostras de algumas das classes da base de dados VGGFace2 [7].

3.1.2 Places365 - Standard

Places365 - Standard é um subconjunto da base Places [96], contendo 2.168.460 imagens de 365 cenas diferentes, sendo 1.803.460 delas para treinamento, 36.500 para validação e 328.500 para teste. Também, estão disponíveis versões desta base com imagens em seus tamanhos originais (em que os conjuntos de treinamento, validação e teste possuem 105 GB, 2,1 GB e 19 GB, respectivamente) ou redimensionadas para 256×256 px (com conjuntos de treinamento, validação e teste possuindo 24 GB, 501 MB e 4,4 GB, nesta ordem), todas



(a) Exemplo 1: quarto. (b) Exemplo 2: quarto. (c) Exemplo 3: quarto.

(d) Exemplo 4: cafeteria. (e) Exemplo 5: cafeteria. (f) Exemplo 6: cafeteria.

(g) Exemplo 7: campo selvagem. (h) Exemplo 8: campo selvagem. (i) Exemplo 9: campo selvagem.

Figura 3.2: Amostras de algumas das classes da base de dados Places365 - Standard [96].

no formato JPEG. Além disso, cada imagem representa apenas uma classe, o conjunto de treinamento possui entre 3.068 e 5.000 amostras por classe e os conjuntos de validação e teste possuem 100 e 900 imagens por classe, respectivamente. Na Figura 3.2, pode-se observar exemplos de categorias presentes na base Places365 - Standard [96], em que se percebe a diversidade de características que compõem um mesmo tipo de cena.

3.1.3 Digipathos Embrapa

A base Digipathos Embrapa [5, 6] reúne imagens de 21 espécies de plantas acometidas por 171 doenças e outras desordens, coletadas ao longo de quatro anos. Ela é composta por duas outras bases de dados, PDDB e XDB, em que a primeira é a base originalmente construída, possuindo 2.337 imagens, e a segunda é o resultado de uma superamostragem realizada sobre a primeira, em que foram destacadas as regiões de lesões e sintomas presentes nas folhas das plantas, obtendo-se um total de 46.101 amostras.

As imagens presentes nesta base possuem resoluções que variam de 1 a 24 $MegaPixels$, todas estão no formato JPEG e a base possui um tamanho de, aproximadamente, 8,3 GB. Cerca de 40% das amostras presentes na base PDDB foram obtidas em campo e cada classe dessa base possui entre 1 e 88 imagens. Por sua vez, para a geração do conjunto XDB, todas as imagens utilizadas tiveram seus planos de fundo removidos e apenas as imagens contendo sintomas em folhas foram empregadas. Como resultado, a base XDB possui apenas 93 das 171 categorias originais e a quantidade de amostras por classe varia de 1 a 3.791. Na Figura 3.3, pode-se observar exemplos de algumas das classes presentes na base Digipathos Embrapa [5, 6], apresentando diferentes sintomas e categorias de cada um dos subconjuntos descritos.

3.1.4 COVIDx

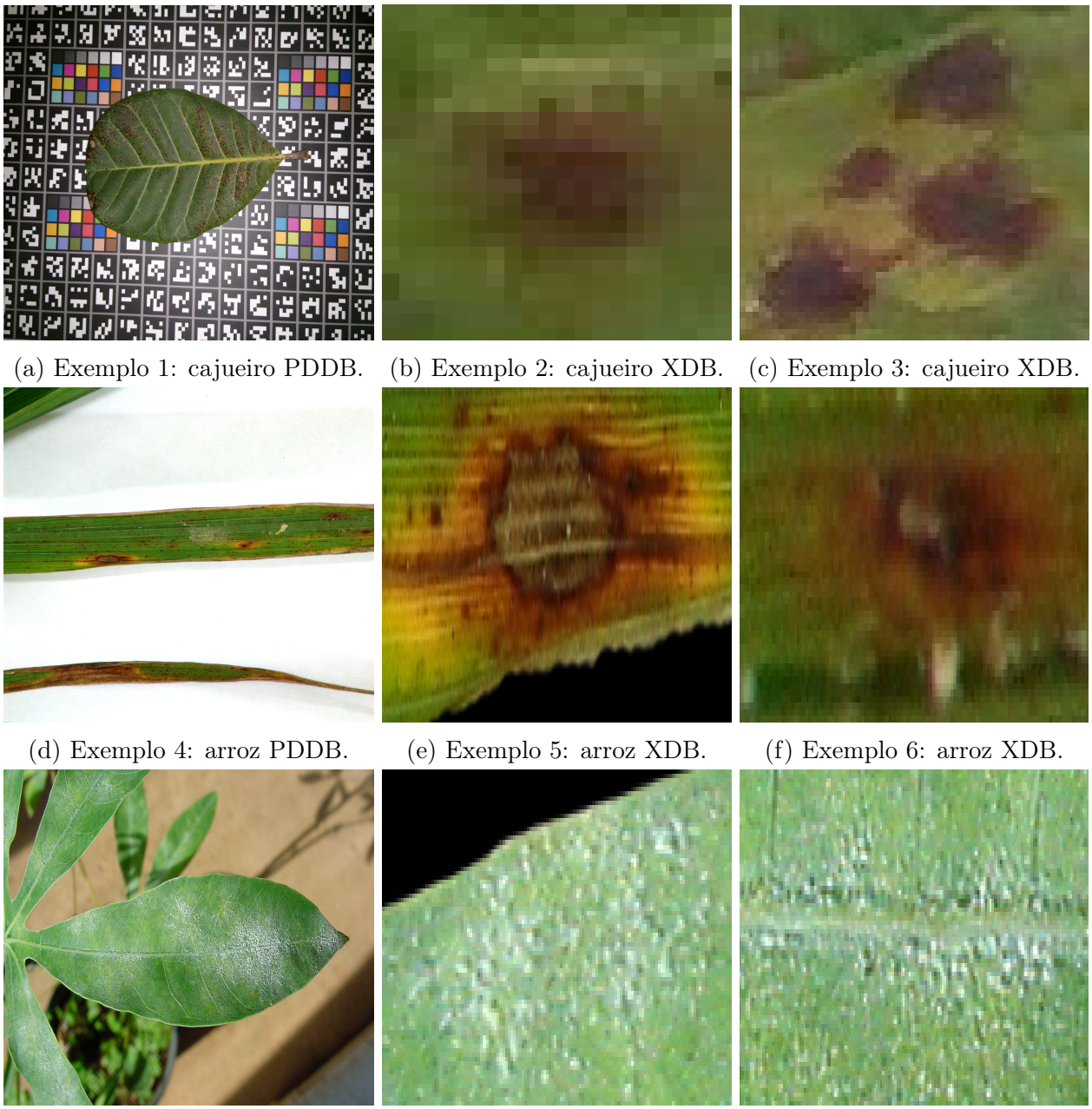
Por último, COVIDx [83] é uma base de dados com 15.514 imagens de radiografias de tórax de 14.002 pacientes, que são utilizadas para se determinar se um paciente possui o novo coronavírus SARS-CoV-2 (COVID-19), se tem algum tipo de pneumonia ou se não possui nenhum desses problemas (normal). Cada imagem possui apenas uma classe, as imagens foram obtidas a partir de outros cinco repositórios públicos que estão continuamente crescendo, e seus formatos variam entre JPEG, PNG (*Portable Network Graphics*) e DICOM (*Digital Imaging and Communications in Medicine*). Além disso, o conjunto de treinamento é composto por 13.935 amostras, o de teste por 1.579 amostras, e eles possuem tamanhos próximos de 5,8 GB e 774,9 MB, respectivamente. Na Figura 3.4, pode-se observar exemplos de cada uma das classes presentes nesta base, obtidos a partir dos diferentes repositórios que a compõem.

3.1.5 Resumo das Bases

Nas Tabelas 3.1 e 3.2, apresenta-se algumas estatísticas sobre a distribuição das amostras para as três primeiras bases e para a COVIDx [83], respectivamente.

3.2 Recursos Computacionais

A implementação deste projeto será realizada em linguagem de programação Python, uma vez que há um grande número de ferramentas desenvolvidas com suporte para essa linguagem e com boa documentação. O projeto utilizará bibliotecas científicas, numéricas, de redes



(a) Exemplo 1: cajueiro PDDB. (b) Exemplo 2: cajueiro XDB. (c) Exemplo 3: cajueiro XDB.

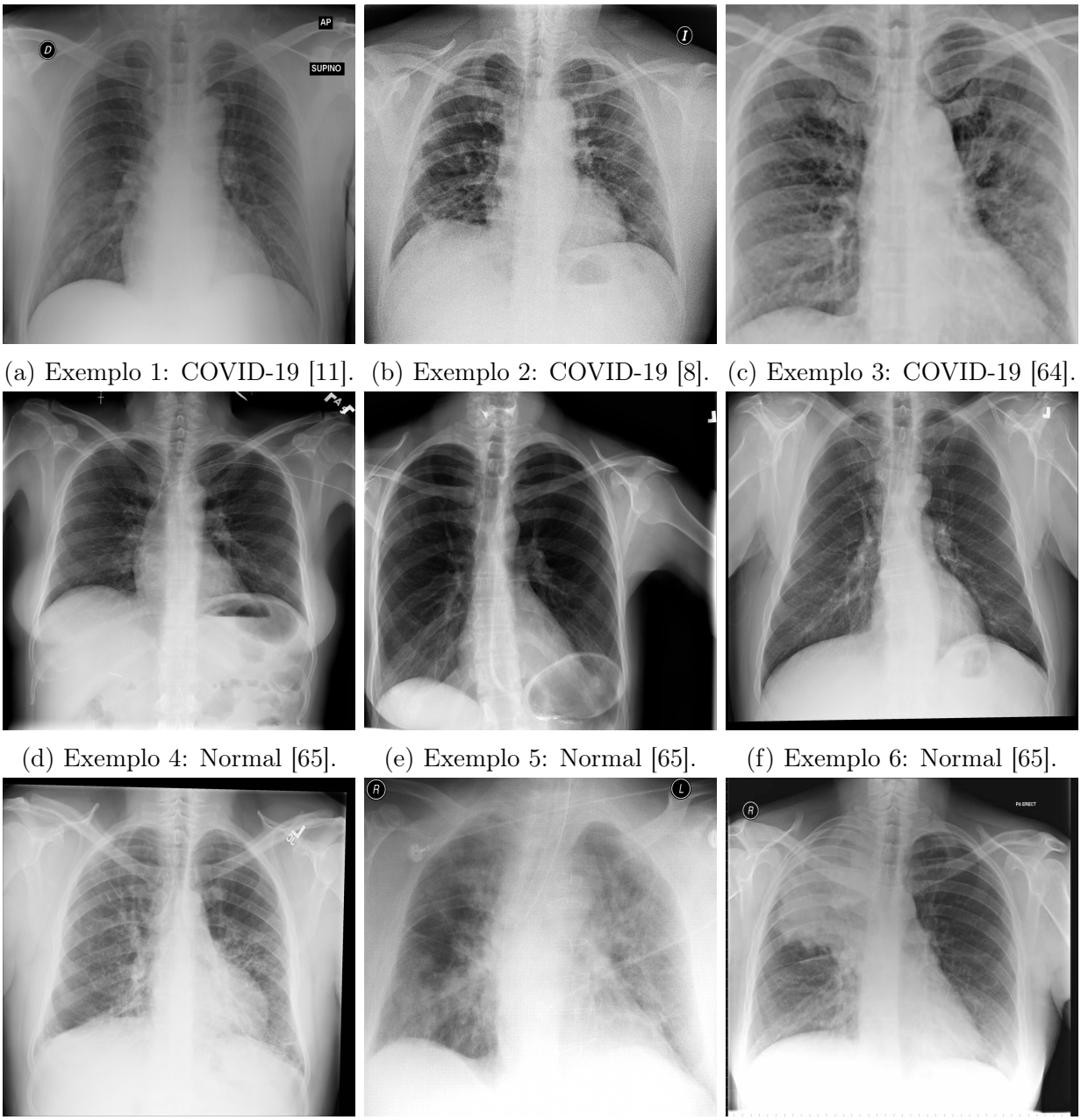
(d) Exemplo 4: arroz PDDB. (e) Exemplo 5: arroz XDB. (f) Exemplo 6: arroz XDB.

(g) Exemplo 7: mandioca PDDB. (h) Exemplo 8: mandioca XDB. (i) Exemplo 9: mandioca XDB.

Figura 3.3: Amostras de algumas das espécies e classes da base Digipathos Embrapa [5, 6].

neurais, de apresentação de gráficos e imagens, entre outras. Algumas biblotecas que podem ser destacadas são: NumPy[1], Pandas[2], scikit-learn[3], OpenCV[4], TensorFlow[5], Keras[6],

[1] `https://numpy.org/`
[2] `https://pandas.pydata.org/`
[3] `https://scikit-learn.org/`
[4] `https://opencv.org/`
[5] `https://www.tensorflow.org/`
[6] `https://keras.io/`



(a) Exemplo 1: COVID-19 [11]. (b) Exemplo 2: COVID-19 [8]. (c) Exemplo 3: COVID-19 [64].

(d) Exemplo 4: Normal [65]. (e) Exemplo 5: Normal [65]. (f) Exemplo 6: Normal [65].

(g) Exemplo 7: Pneumonia [65]. (h) Exemplo 8: Pneumonia [11]. (i) Exemplo 9: Pneumonia [11].

Figura 3.4: Amostras de algumas das classes da base de dados COVIDx [83].

Matplotlib[7] e PyTorch[8].

Os experimentos deste projeto serão realizados em dois ambientes: no Laboratório de Informática Visual (LIV) do Instituto de Computação (IC - Unicamp), em uma máquina equipada com processador Intel i7-3770, 3.50 GHz, e com uma GPU NVIDIA TITAN V, com 5120 núcleos e 12 GB de memória; e na plataforma Google Colab[9], que fornece recursos

[7] `https://matplotlib.org/`
[8] `https://pytorch.org/`
[9] `https://colab.research.google.com/`

Tabela 3.1: Estatísticas sobre as amostras para as bases de dados.

				Quantidade de amostras por classe		
Base de dados	Subconjunto	Número de classes	Total de amostras	Mínimo	Média	Máximo
Digipathos Embrapa [5, 6]	PDDB	171	2.337	1	13	88
	XDB	93	46.101	1	495	3.791
Places365 - Standard [96]	Treinamento	365	1.803.460	3.068	4.940	5.000
	Validação		36.500	100	100	100
	Teste		328.500	900	900	900
VGGFaces2 [7]	Treinamento	8.631	3.141.890	87	364	843
	Teste	500	169.396	98	338	761

Tabela 3.2: Estatísticas sobre a distribuição de amostras da base de dados COVIDx.

Base de dados	Subconjunto	Classe	Total de amostras
COVIDx [83]	Treinamento	COVID-19	505
		Pneumonia	5.464
		Normal	7.966
	Teste	COVID-19	100
		Pneumonia	594
		Normal	885

de computação em nuvem gratuitos, tais como CPU, GPU, TPU, memória e armazenamento.

Capítulo 4

Métodos

Neste capítulo, descrevemos a metodologia inicial, que será utilizada na investigação de técnicas de aumentação de dados para classificação de imagens, assim como as métricas de avaliação.

4.1 Metodologia

Nesta seção, apresentamos a metodologia proposta divida em três partes: treinamento, validação e teste; otimização de hiperparâmetros e número de épocas; e protocolo experimental. Na Figura 4.1, estão ilustradas as principais etapas de experimentação.

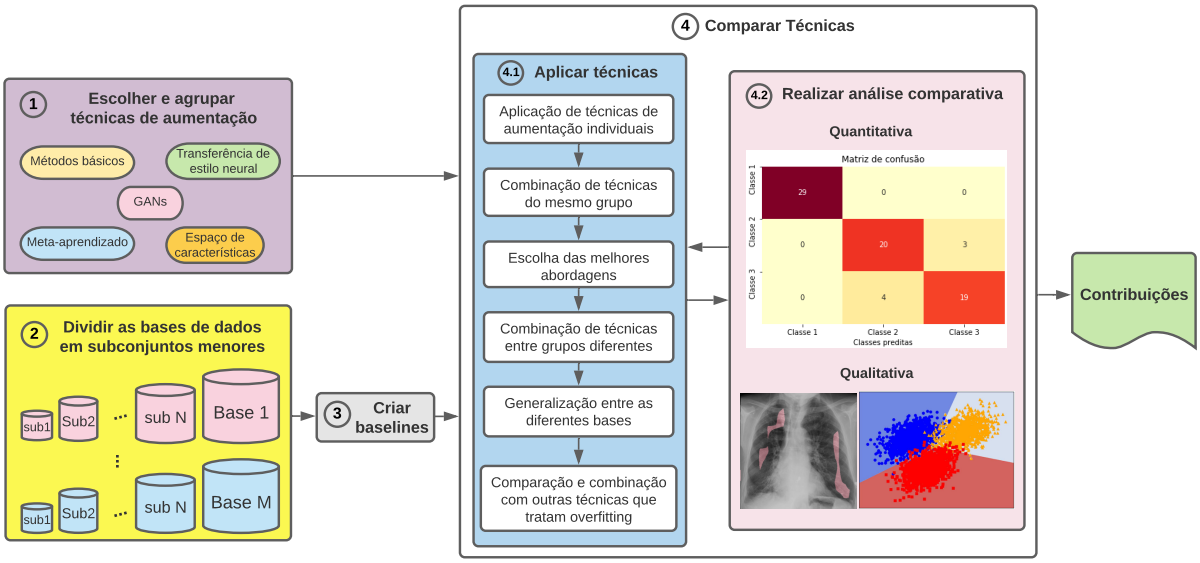


Figura 4.1: Diagrama com as principais etapas que compõem a metodologia deste projeto.

4.1.1 Treinamento, Validação e Teste

Neste projeto, investigaremos, inicialmente, quatro bases de dados descritas na Seção 3.1, algumas das quais fornecem conjuntos separados de treinamento, validação e teste, enquanto outras não. Tendo isso em vista, durante as etapas de treinamento e validação, utilizaremos tanto os métodos de validação cruzada padrão (*holdout*), assim como a técnica de *k-fold*, de maneira a escolhermos os melhores modelos e seus respectivos hiperparâmetros.

Uma vez encontradas as melhores configurações de arquiteturas e hiperparâmetros para cada base, dois modelos serão empregados para avaliação sobre o conjunto de teste, um deles treinado apenas sobre os dados de treinamento e o outro sobre a junção dos conjuntos de treinamento e validação.

Como as bases de dados VGGFace2 [7] e COVIDx [83] possuem conjuntos de treinamento e teste separados, mas não há conjuntos de validação, utilizaremos o processo de validação cruzada *k-fold* sobre os dados de treinamento. Além disso, apenas 50 imagens de cada classe do conjunto de teste da base VGGFace2 serão testadas, como feito por Cao et al. [7]. Já a base de dados Places365-Standard [96] fornece os três conjuntos separados, de modo que utilizaremos a validação cruzada padrão. Por fim, para a base de dados Digipathos Embrapa [5], seguiremos os passos realizados por Barbedo et al. [5], em que 80% das amostras são utilizadas para treinamento e validação, e 20 % são usadas para teste, com um processo de validação cruzada 10-*fold*.

Todas as imagens serão pré-processadas como descrito nos artigos de suas respectivas bases de dados [5, 7, 83, 96], o que inclui seu redimensionamento e a normalização dos valores dos *pixels*. Também, diversos classificadores serão experimentados, sendo um conjunto inicial composto pelos modelos que obtiveram os melhores desempenhos reportados nos artigos: SE-ResNet-50 na base VGGFace2; VGG16 e ResNet-152 na base Places365-Standard; GoogLeNet (e arquiteturas *Inception* mais recentes) no conjunto Digipathos Embrapa; e ResNet-50 e COVID-Net no conjunto de dados COVIDx.

4.1.2 Otimização de Hiperparâmetros e Número de Épocas

Os hiperparâmetros de cada classificador e método de aumentação de dados serão estudados e buscas em grade (*Grid Search*) serão realizadas para a escolha dos melhores valores. Com relação aos métodos de aumentação de dados, os hiperparâmetros dizem respeito tanto às redes neurais (para as técnicas baseadas em aprendizado profundo) como ao próprio espaço de busca de aumentações (como o número de graus que uma imagem deve ser rotacionada). Neste último caso, determinadas suposições sobre os valores desses hiperparâmetros poderão ser empregadas para a diminuição do espaço de buscas.

Para a escolha do melhor número de épocas, para a diminuição do tempo de treinamento e para a prevenção de *overfitting* devido a um treinamento excessivo, diversos critérios de parada antecipada serão experimentados, tais como a acurácia e o custo de validação, para diferentes valores de paciência (número de épocas sem melhoria das métricas).

4.1.3 Protocolo Experimental

Os experimentos serão conduzidos em quatro etapas, conforme descritas a seguir:

Escolha e Agrupamento das Técnicas de Aumentação (1)

Para a condução dos experimentos, devemos selecionar as técnicas de aumentação a serem utilizadas. Inicialmente, trabalharemos com os métodos apresentados no Capítulo 2, conforme a disponibilidade dos códigos e as restrições para sua implementação. Todavia, ao longo do desenvolvimento deste projeto, outros métodos poderão ser adicionados ou alguns dos já selecionados poderão ser excluídos.

Uma vez selecionadas as técnicas de aumentação de dados, elas serão agrupadas em cinco categorias: métodos básicos, aumentação no espaço de características, aumentação baseada em GANs, transferência de estilo neural e técnicas baseadas em meta-aprendizado. Ademais, cada categoria ainda poderá ser subdividida, como no caso dos métodos básicos, em que poderíamos separar transformações geométricas da adição de ruídos e da combinação de imagens.

Divisão das Bases de Dados em Subconjuntos de Diferentes Tamanhos (2)

Cada base de dados terá seu conjunto de treinamento dividido em subconjuntos de diferentes tamanhos, de modo que investigações sobre o efeito da aumentação de dados sobre bases com diferentes quantidades de amostras possam ser realizadas [2].

Com a criação de conjuntos de variados tamanhos, poderemos simular cenários que possuem uma pequena quantidade de dados de treinamento, os quais são os mais propensos à geração de modelos com *overfitting*. Também, ao experimentarmos cenários com um número cada vez maior de amostras, poderemos avaliar como os métodos de aumentação influenciam o desempenho dos modelos, não somente devido à adição de mais dados, mas também pela diversidade que é inserida com essas novas amostras, melhorando a qualidade do conjunto.

Podemos observar que essas são questões que podem ser investigadas, inclusive, nos conjuntos originais, os quais possuem diferentes tamanhos e sofrem com desbalanceamento entre as classes, conforme apresentado nas Tabelas 3.1 e 3.2.

Criação de Baselines (3)

Para observarmos as influências dos métodos de aumentação, serão realizados experimentos com cada subconjunto sem aumentação de dados (*baselines*), inclusive o original, e com aumentação.

Comparação das Técnicas de Aumentação (4)

Para a comparação das técnicas, dois processos serão realizados concomitantemente: a aplicação dos métodos de aumentação (4.1) e a análise de sua influência sobre o desempenho dos classificadores e sobre a qualidade do conjunto de dados (4.2). A aplicação das abordagens de aumentação de dados podem ser divididas em cinco etapas:

- Cada método será aplicado individualmente e comparações serão feitas entre os métodos pertencentes ao mesmo grupo.
- Técnicas de um mesmo grupo serão combinadas e comparadas entre si.
- As melhores abordagens encontradas em cada categoria serão utilizadas para novas combinações, desta vez, entre as diferentes categorias.
- Como as melhores técnicas podem variar de acordo com a base de dados, realizaremos experimentos com as melhores abordagens, encontradas em cada base, sobre todas as outras bases, de maneira a investigarmos o potencial de generalização dessas abordagens.
- Além disso, analisaremos como as técnicas de aumentação interagem com outras abordagens que tratam o *overfitting*, tais como regularização *dropout* [75] e o pré-treinamento em bases de dados semelhantes [28]. Para isso, *baselines* serão construídos apenas com essas técnicas alternativas e compararemos seus desempenhos com os de modelos treinados apenas com a aumentação de dados e com a combinação entre essas abordagens.

Serão experimentadas tanto aumentações *online* como *offline*, de acordo com as limitações de cada método e com os recursos computacionais disponíveis.

Para a análise comparativa entre as técnicas investigadas, utilizaremos duas abordagens, uma quantitativa e a outra qualitativa. Na análise quantitativa, a influência das técnicas de aumentação sobre os desempenhos dos modelos em uma determinada base, e sua capacidade de generalização para outros conjuntos de dados, serão analisadas de acordo com as métricas descritas na Seção 4.2, observando-se quais métricas são utilizadas em cada base.

Já na abordagem qualitativa, analisaremos que informações os algoritmos consideram para a tomada de decisões [83], de maneira a prevenir a utilização de informações enviesadas e indicadores visuais irrelevantes, que levam os modelos a um aprendizado incorreto. Além disso, observaremos como a aumentação altera o *manifold* dos conjuntos de dados [98] e como isso se relaciona com os desempenhos dos modelos.

4.2 Métricas de Avaliação de Desempenho

Os modelos de aprendizagem de máquina podem ser empregados para diversos tipos de tarefa, como classificação, regressão e clusterização. Independentemente do tipo de tarefa para o qual eles são construídos, ferramentas que mensuram o seu desempenho devem ser utilizadas, as quais são chamadas de métricas de avaliação de desempenho. Nesta seção, apresentaremos as principais métricas que empregaremos para a avaliação dos modelos propostos neste projeto.

4.2.1 Sensibilidade

Uma das métricas de avaliação utilizadas em problemas de classificação é a sensibilidade, também chamada de revocação (*recall*). Ela é definida como a razão entre a quantidade de

predições positivas verdadeiras e o número total de amostras positivas, conforme mostrada na Equação 4.1, em que TP (*True Positive*) equivale à quantidade de predições positivas verdadeiras, TN (*True Negative*) indica o número de predições negativas verdadeiras, FP (*False Positive*) refere-se ao número de falsos positivos e FN (*False Negative*) equivale ao número de falsos negativos.

$$\text{Sensibilidade} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.1}$$

A sensibilidade pode ser interpretada como um indicativo da quantidade de amostras positivas que foram preditas como positivas, sofrendo forte influência dos falsos negativos, de modo que, quanto maior é o número de falsos negativos, menor é a sensibilidade. Por exemplo, pode-se pensar que ela diz quantos dos pacientes que possuem uma determinada doença são diagnosticados com essa doença.

4.2.2 Acurácia

A acurácia é uma das métricas de avaliação mais utilizadas para a determinação do desempenho de modelos em problemas de classificação. Ela é dada pela razão entre o número de predições corretas e a quantidade total de predições, conforme apresentada na Equação 4.2.

$$\text{Acurácia} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4.2}$$

Um detalhe importante é que a acurácia assume que o conjunto de dados é balanceado, ou seja, que todas as classes possuem a mesma quantidade de amostras, falhando em representar o desempenho real dos modelos quando isso não acontece. Dessa maneira, quando a base de dados é desbalanceada, outras métricas de avaliação devem ser utilizadas. Uma delas é a acurácia balanceada, apresentada na Equação 4.3, em que N é o número de classes e c_i corresponde à classe i. Em outras palavras, a acurácia balanceada é a média aritmética das sensibilidades de todas as classes.

$$\text{Acurácia Balanceada} = \frac{1}{N} \sum_{i=1}^{N} \text{Sensibilidade}(c_i) \tag{4.3}$$

Além disso, em alguns problemas de classificação, o modelo tem como saída não apenas o rótulo da amostra dada como entrada, mas um vetor de probabilidades em que cada elemento está associado a uma classe diferente da base de dados e indica a probabilidade dessa classe ser o rótulo correto para a amostra. Logo, introduzimos mais duas métricas: a acurácia *top*-1 e a acurácia *top*-5. A primeira indica a proporção de predições em que a classe com maior probabilidade é a classe verdadeira da amostra. Por sua vez, a acurácia *top*-5 representa a porcentagem das predições em que a classe verdadeira está entre os cinco rótulos preditos com maior probabilidade.

4.2.3 Precisão

A precisão, também chamada de valor preditivo positivo (*Positive Predictive Value* - PPV), é dada pela razão entre o número de predições positivas verdadeiras e a quantidade total de predições positivas, como pode-se observar na Equação 4.4.

$$\text{Precisão} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.4}$$

Essa métrica indica quantas das predições ditas como positivas são realmente positivas, sendo bastante influenciada pela quantidade de falsos positivos, de forma que, quanto maior é o número de falsos positivos, menor é a precisão. Como exemplo, ela indica quantos dos pacientes diagnosticados com uma determinada doença realmente possuem essa doença.

A sensibilidade, a acurácia e a precisão podem ser obtidas a partir das chamadas matrizes de confusão. Uma matriz de confusão é uma matriz quadrada, em que as linhas representam as classes verdadeiras das amostras, enquanto as colunas representam as classes preditas. Sempre que uma nova amostra é classificada, incrementamos a posição da tabela correspondente à linha de sua categoria verdadeira e à coluna da categoria predita.

4.2.4 Outras Métricas

Em alguns outros domínios de reconhecimento de padrões, outras métricas podem ser empregadas na avaliação dos resultados: a Taxa de Identificação de Positivo Verdadeiro (*True Positive Identification Rate* - TPIR), a Taxa de Identificação de Falso Negativo (*False Negative Identification Rate* - FNIR), a Taxa de Identificação de Falso Positivo (*False Positive Identification Rate* - FPIR), o *Rank* de Identificação, a Taxa de Aceitação Verdadeira (*True Acceptance Rate* - TAR) e a Taxa de Aceitação Falsa (*False Acceptance Rate* - FAR).

Quando a tarefa a ser desempenhada nesse domínio é a de identificação, dois tipos de conjuntos de referência podem ser utilizados, chamados de conjunto fechado (*closed-set*), em que todos os indivíduos a serem testados pertencem à base em que as identidades serão buscadas, e de conjunto aberto (*open-set*), em que algumas das amostras testadas são de indivíduos não pertencentes à base de referência. Dessa maneira, algumas das métricas citadas podem ser definidas como:

- TPIR: é a proporção de consultas em que a identidade correta da amostra está entre as k identidades retornadas e possui uma grau de semelhança maior ou igual a um determinado limiar T. Também pode ser chamada de taxa de acerto.

- FNIR: é a proporção de consultas em que a identidade correta da amostra não está entre as k identidades retornadas ou o grau de semelhança está abaixo de um determinado limiar T. Também pode ser chamada de taxa de perda e é definida como $1 - \text{TPIR}$.

- FPIR: é a proporção de consultas em que a amostra não pertence à base de referência, mas uma ou mais identidades são retornadas, com um grau de semelhança maior do que um determinado limiar T. Ela também pode ser chamada de taxa de alarme falso.

- *Rank* de Identificação: é definido como o menor valor de k para o qual a identidade correta da amostra consultada está entre as k identidades retornadas.

Por outro lado, quando a tarefa a ser executada é a verificação, as métricas FAR e TAR são geralmente empregadas, podendo ser definidas como:

- FAR: é a proporção de consultas em que as duas amostras são consideradas como pertencentes à mesma identidade quando, na verdade, elas não pertencem.
- TAR: é a proporção de consultas em que considera-se que as duas amostras pertencem à mesma identidade quando elas realmente pertencem.

Além dessas métricas, curvas de desempenho também são geralmente empregadas, sendo três delas a Troca Erro Detecção (*Detection Error Tradeoff* - DET) e a Característica de Correspondência Cumulativa (*Cumulative Match Characteristic* - CMC), para a tarefa de identificação, e a ROC (*Receiver Operating Characteristic*), para a tarefa de verificação.

A ROC é uma curva de probabilidades geralmente empregada na avaliação de modelos de classificação, que facilita a escolha de determinadas características desses modelos, de maneira a se atingir melhores resultados. No domínio de verificação de faces, a ROC pode ser construída como uma curva no plano cartesiano, em que no eixo horizontal estão distribuídos os valores de FAR, enquanto no eixo vertical está representada a métrica TAR. Assim, quanto maior for o valor obtido nesta curva, melhor é o desempenho do que está sendo medido.

A DET é um tipo de curva ROC em que ambos os eixos representam taxas de erro, sendo FPIR a taxa correspondente ao eixo horizontal e FNIR a do vertical, de maneira que quanto menor for o valor mensurado na curva, melhor é o desempenho do que está sendo medido.

Por sua vez, a curva CMC representa a relação entre a taxa TPIR e o *rank* de identificação. Em outras palavras, para cada valor k do *rank*, distribuído no eixo horizontal, existe um valor de TPIR correspondente, no eixo vertical, que indica a probabilidade de a identidade correta da amostra estar entre as k identidades do *rank*. Dessa maneira, quanto maior for o valor obtido na curva, melhor é o desempenho do que está sendo medido.

Capítulo 5

Resultados Preliminares

Neste capítulo, apresentamos alguns resultados preliminares obtidos sobre a base de dados COVIDx, descrita na Seção 3.1.

Como classificador, escolhemos a ResNet-50, que foi uma das redes utilizadas no trabalho de Wang et al. [83]. As configurações adotadas nas etapas de treinamento e validação foram: otimizador SGD (*Stochastic Gradient Descent*), taxa de aprendizado 0,001, momentum Nesterov de 0,9, Entropia Cruzada Categórica como função de perda e 100 épocas como duração limite da etapa de treinamento. Como a base já possui conjuntos de teste e treinamento separados, apenas este último foi dividido em 20% para validação, e uma validação cruzada padrão foi empregada. Além disso, todos os arquivos com as separações dos conjuntos foram previamente gerados e armazenados, de maneira que os mesmos dados foram utilizados em todos os experimentos. Neste momento, utilizamos apenas a precisão, a sensibilidade e a acurácia balanceada, pois são métricas utilizadas no trabalho de referência da base COVIDx [83]. Após o treinamento, predições foram realizadas sobre os conjuntos de treinamento, validação e teste.

Visando analisar como o uso de aumentação de dados influencia o desempenho de classificadores sobre a base COVIDx, duas abordagens de experimentação foram empregadas, uma sem e a outra com aumentação de dados tradicional, e cada experimento foi executado apenas uma vez. A aumentação de dados tradicional foi aplicada de forma *online*, consistindo nas seguintes operações, empregadas por Wang et al. [83]: rotação, translação, alteração de brilho, escala e espelhamento horizontal. Para os experimentos, os valores dessas transformações foram escolhidos aleatoriamente. Com o objetivo de investigarmos como o tamanho do conjunto de dados influencia o desempenho dos modelos de aprendizado e como essa característica interage com as abordagens de experimentação descritas, em cada experimento simulamos três cenários de restrições sobre a quantidade de amostras, chamados de *COVID*, *Balanceado* e *Todas as Classes*. Para cada cenário, criamos 10 subconjuntos de treinamento, incrementando a quantidade de dados através da aplicação das percentagens 10% a 100%, em passos de 10%, sobre uma ou mais classes, de acordo com o cenário a ser simulado, como descrito a seguir:

- *COVID-19*: suponha que existam algumas doenças já bem conhecidas pela humanidade, sobre as quais possuímos informações suficientes. Todavia, uma nova doença

surge e, inicialmente, temos poucos dados sobre ela, os quais vão aumentando ao longo do tempo. Simulamos este cenário fixando a quantidade de amostras das classes que representam as doenças já conhecidas (no caso da base COVIDx, as categorias *Normal* e *Pneumonia*) e incrementando a quantidade de amostras da nova doença (classe *COVID-19*). Os conjuntos gerados possuem 10.677, 10.717, 10.757, 10.797, 10.837, 10.877, 10.917, 10.957, 10.997 e 11.037 amostras, respectivamente.

- *Balanceado*: suponha agora que surjam várias doenças novas e que temos, aproximadamente, a mesma pequena quantidade de dados sobre elas, que também aumenta ao longo do tempo. Para simular este cenário, amostramos o mesmo número de imagens para todas as classes da base COVIDx, o qual é determinado pela aplicação das percentagens sobre a classe minoritária. Os conjuntos resultantes possuem 120, 240, 360, 480, 600, 720, 840, 960, 1.080 e 1.200 amostras.

- *Todas as Classes*: neste cenário, supomos que existem doenças sobre as quais possuímos algumas informações. Então, surge uma nova doença sobre a qual temos ainda menos dados. No entanto, com o passar do tempo, conseguimos mais informações sobre todas elas. Para simulá-lo, aplicamos a mesma percentagem sobre cada classe. Assim, se a classe *COVID-19* aumenta em 10%, as classes *Normal* e *Pneumonia* também aumentarão. Os tamanhos dos conjuntos gerados neste cenário são 1.104, 2.207, 3.311, 4.415, 5.518, 6.622, 7.726, 8.829, 9.933 e 11.037 amostras.

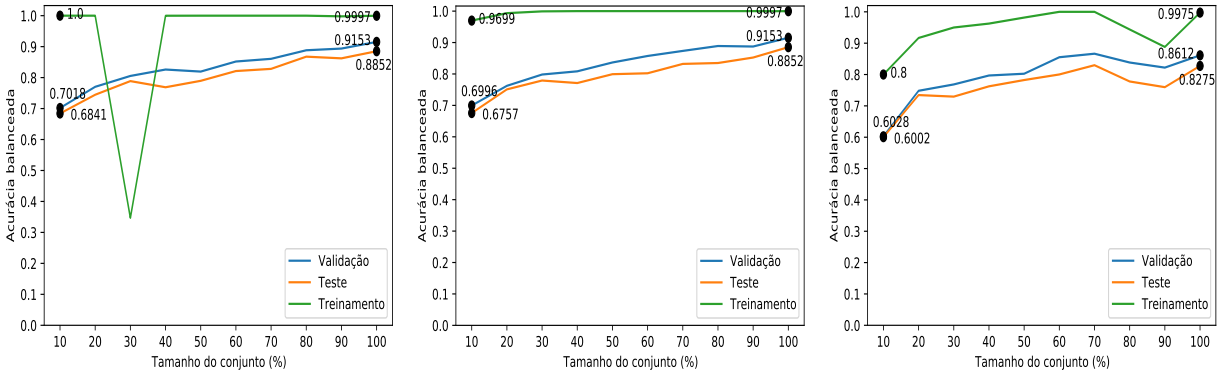
A seguir, discutimos os resultados de cada experimento.

5.1 Resultados e Discussões

As Figuras 5.1 e 5.2 apresentam os gráficos de acurácia balanceada, de acordo com o tamanho do subconjunto de treinamento, para os experimentos realizados com as imagens originais e com aumentações de dados tradicionais, respectivamente. Em ambas as figuras, vemos que as acurácias obtidas sobre o conjunto de treinamento são maiores do que aquelas obtidas sobre os conjuntos de validação e teste, como esperado. Também, observamos que os valores de validação e teste são próximos, principalmente nos experimentos com aumentação de dados, indicando que as distribuições desses conjuntos são semelhantes, o que é desejado.

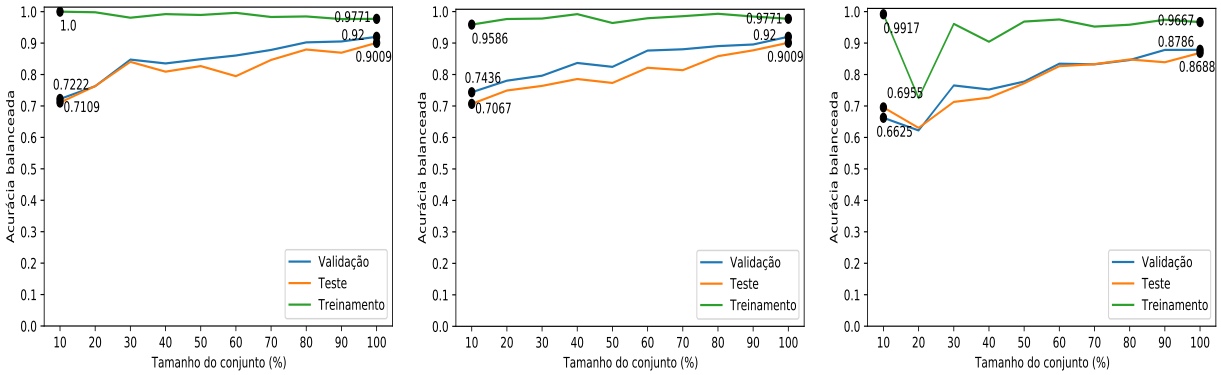
Apesar disso, podemos observar que, nos experimentos sem aumentação de dados, as acurácias de treinamento atingem valores próximos ao máximo, para a maioria dos tamanhos, principalmente nos cenários *Todas as Classes* e *COVID*, ao passo que esses valores mostram-se mais baixos quando os modelos são treinados com amostras aumentadas. Ademais, em ambas as abordagens de experimentação, as acurácias de validação e teste crescem ao incrementarmos o tamanho dos conjuntos de treinamento, e os valores obtidos ao empregarmos aumentação de dados são maiores do que quando não a utilizamos.

Logo, essas observações indicam que as técnicas de aumentação de dados adotadas reduziram o *overfitting* dos modelos, diminuindo seu desempenho sobre o conjunto de treinamento, mas aumentando-o sobre os conjuntos de validação e teste.



(a) Cenário *Todas as Classes*. (b) Cenário *COVID*. (c) Cenário *Balanceado*.

Figura 5.1: Acurácias balanceadas para experimentos sem aumentação de dados.



(a) Cenário *Todas as Classes*. (b) Cenário *COVID*. (c) Cenário *Balanceado*.

Figura 5.2: Acurácias balanceadas para experimentos com aumentação de dados.

As Tabelas 5.1 e 5.2 apresentam os valores de precisão e sensibilidade de teste para os experimentos realizados sem aumentação de dados. Observamos que, nos cenários *Todas as Classes* e *COVID*, as precisões da classe *Normal* são maiores do que as da classe *COVID-19*, que são maiores do que as da categoria *Pneumonia*, em grande parte dos tamanhos dos conjuntos de treinamento. Porém, ao analisarmos a sensibilidade, percebemos que os valores da categoria *COVID-19* são muito menores do que os das outras duas.

Isso demonstra que, apesar de uma elevada parcela das amostras que o modelo prediz como *COVID-19* pertencer a esta classe, apenas uma pequena proporção dos casos de *COVID-19* são identificados. Ademais, tendo em vista os melhores desempenhos sobre a categoria *Normal*, as menores precisões sobre a classe *Pneumonia* e a maior semelhança que é esperada entre as imagens desta e as de *COVID-19*, é provável que muitas das amostras que contribuem para a baixa sensibilidade da classe *COVID-19* estejam sendo erroneamente preditas como *Pneumonia*. Isso é confirmado pelas matrizes de confusão das Figuras 5.3a a 5.3d e das Figuras 5.4a a 5.4d, para os experimentos sem e com aumentação de dados, respectivamente.

Ao analisarmos o cenário *Balanceado*, podemos observar que tanto os valores de precisão, como os de sensibilidade, são menores do que os dos outros cenários. Dois fatores podem

Tabela 5.1: Precisão para os experimentos sem aumentação de dados.

Precisão									
Tamanho	Todas as Classes			COVID			Balanceado		
(%)	Normal	Pneumonia	COVID-19	Normal	Pneumonia	COVID-19	Normal	Pneumonia	COVID-19
10	0,9144	0,8171	0,9565	0,9108	0,8476	0,9444	0,8856	0,6684	0,8235
20	0,9311	0,8195	0,9302	0,9313	0,8616	**0,9500**	0,8730	0,7798	0,5091
30	0,9361	0,8657	0,9259	0,9301	0,8860	0,9200	0,8851	0,8160	0,7333
40	0,9233	0,8569	0,9200	0,9344	0,8673	0,9362	0,9051	0,8285	0,6842
50	0,9320	0,8688	0,8947	0,9405	0,8736	0,9455	0,8924	0,8152	0,7722
60	0,9428	0,8838	**0,9667**	0,9323	0,8790	0,9310	0,9295	0,8384	0,6854
70	0,9376	0,8883	0,9531	0,9384	0,8864	0,9000	0,9174	**0,8552**	0,6460
80	0,9420	0,9010	0,9481	0,9478	0,8778	0,9265	0,9106	0,8347	0,8182
90	**0,9506**	0,8792	0,9114	**0,9502**	0,8926	0,9437	0,9050	0,8331	0,8033
100	0,9422	**0,9191**	0,9070	0,9422	**0,9191**	0,9070	**0,9357**	0,8383	**0,8375**

Tabela 5.2: Sensibilidade para os experimentos sem aumentação de dados.

Sensibilidade									
Tamanho	Todas as Classes			COVID			Balanceado		
(%)	Normal	Pneumonia	COVID-19	Normal	Pneumonia	COVID-19	Normal	Pneumonia	COVID-19
10	0,9299	0,9024	0,2200	0,9582	0,8990	0,1700	0,7785	0,8822	0,1400
20	0,9164	0,9175	0,4000	0,9503	0,9226	0,3800	0,8621	0,7811	0,5600
30	0,9435	0,9226	0,5000	**0,9616**	0,9158	0,4600	0,9051	0,8434	0,4400
40	0,9390	0,9074	0,4600	0,9492	0,9242	0,4400	0,9051	0,8620	0,5200
50	0,9446	0,9141	0,5100	0,9469	0,9310	0,5200	0,8904	0,8468	0,6100
60	0,9492	**0,9343**	0,5800	0,9492	0,9175	0,5400	0,9085	0,8822	0,6100
70	0,9503	0,9242	0,6100	0,9469	0,9192	0,6300	0,9040	0,8552	**0,7300**
80	0,9537	0,9192	0,7300	0,9435	0,9310	0,6300	0,9096	0,8838	0,5400
90	0,9356	0,9310	0,7200	0,9492	**0,9377**	0,6700	**0,9153**	0,8737	0,4900
100	**0,9582**	0,9175	**0,7800**	0,9582	0,9175	**0,7800**	0,9051	**0,9074**	0,6700

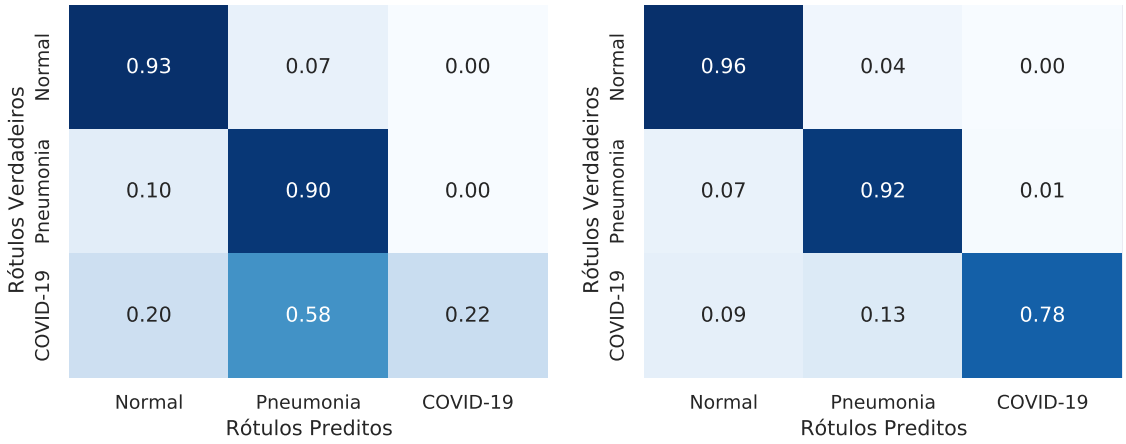
contribuir para isso: a menor quantidade de amostras desse cenário, quando comparada à dos outros; e, uma vez que as categorias estão balanceadas, há uma maior possibilidade delas serem confundidas entre si, ao contrário do que ocorre nos cenários desbalanceados, em que a classe *COVID-19* é a mais prejudicada. Nas Figuras 5.3e e 5.3f, para os experimentos sem aumentação, e nas Figuras 5.4e e 5.4f, para aqueles com aumentação de dados, podemos confirmar essa hipótese.

As Tabelas 5.3 e 5.4 mostram os valores de precisão e sensibilidade de teste para os experimentos realizados com aumentação de dados. Percebemos que todas as classes são beneficiadas pela adoção dessa técnica, principalmente *Pneumonia* e *COVID-19*. Apesar disso, observamos um comportamento semelhante ao encontrado nos experimentos sem aumentação: as precisões da categoria *Pneumonia* apresentam valores menores do que os de *COVID-19*, para a maioria dos tamanhos; as sensibilidades desta última classe são muito menores do que as das outras categorias; e o cenário *Balanceado* possui os menores valores de precisão e sensibilidade.

A partir das análises realizadas nestes resultados preliminares, constatamos que:

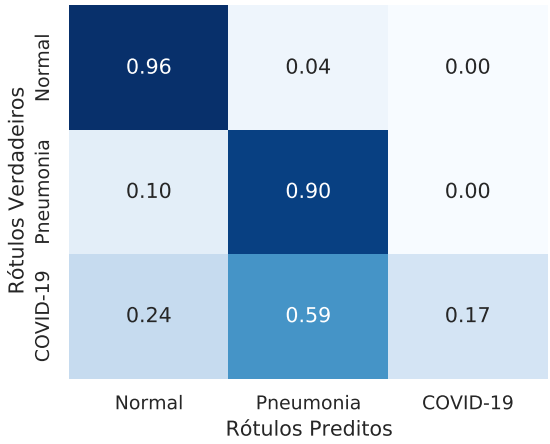
- O desempenho dos classificadores melhora ao incrementarmos o tamanho do conjunto de dados utilizado para treiná-los.

- Apesar disso, apenas a expansão do volume de dados não é suficiente para a construção

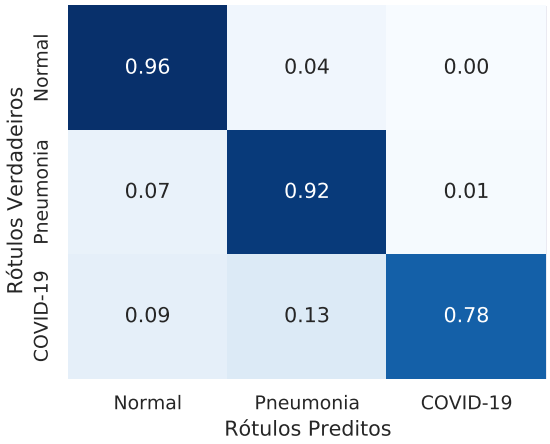


(a) *Todas as Classes*, 10% dos dados.

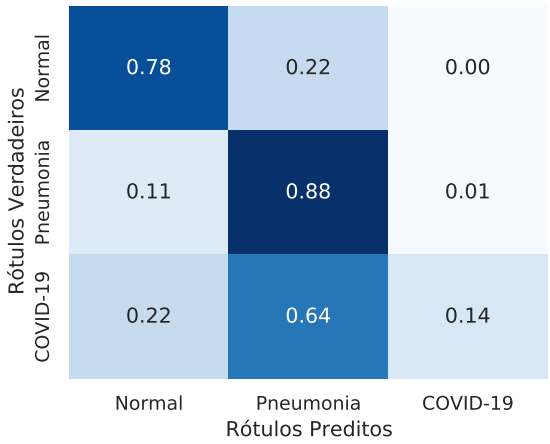
(b) *Todas as Classes*, 100% dos dados.



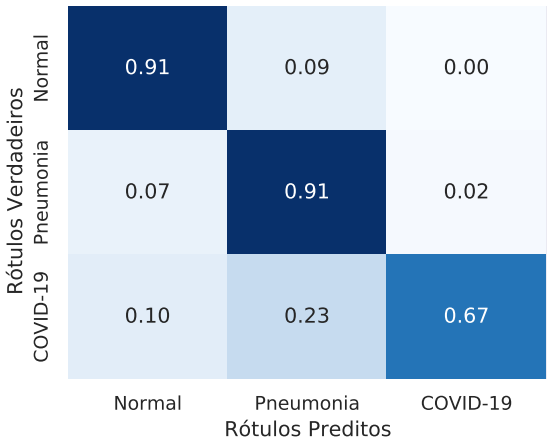
(c) *COVID*, 10% dos dados.



(d) *COVID*, 100% dos dados.



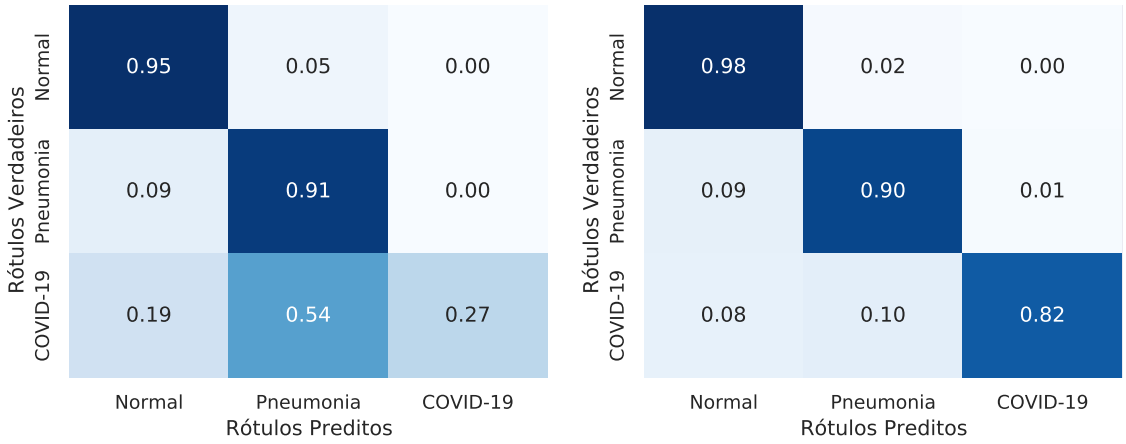
(e) *Balanceado*, 10% dos dados.



(f) *Balanceado*, 100% dos dados.

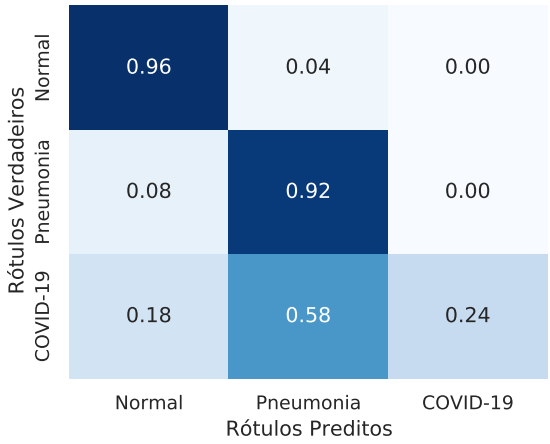
Figura 5.3: Matrizes de confusão sobre o conjunto de testes, sem aumentação de dados.

de bons modelos. Com a adoção de técnicas de aumentação de dados tradicionais, conseguimos reduzir o impacto do *overfitting*, melhorando a acurácia de teste, inclusive

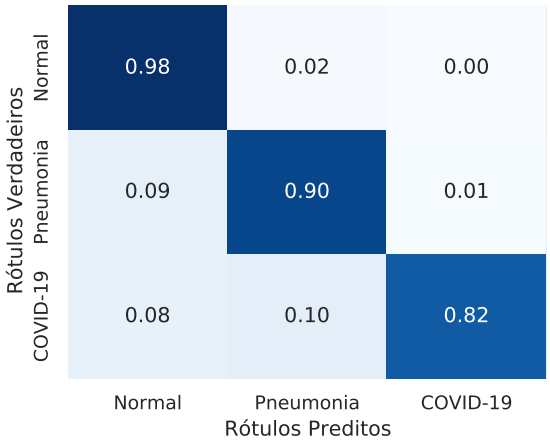


(a) *Todas as Classes*, 10% dos dados.

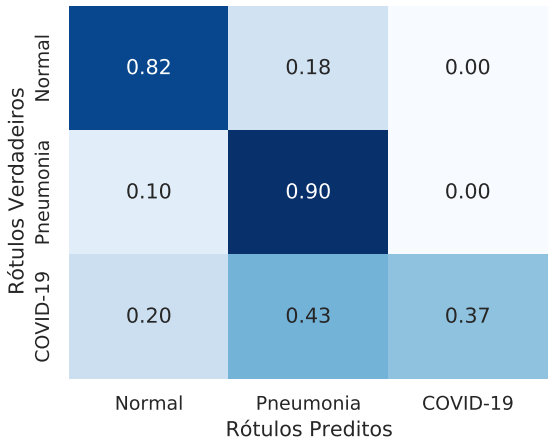
(b) *Todas as Classes*, 100% dos dados.



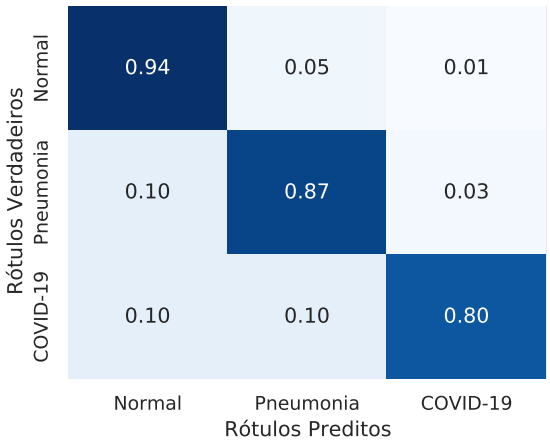
(c) *COVID*, 10% dos dados.



(d) *COVID*, 100% dos dados.



(e) *Balanceado*, 10% dos dados.



(f) *Balanceado*, 100% dos dados.

Figura 5.4: Matrizes de confusão sobre o conjunto de testes, com aumentação de dados.

sobre conjuntos com poucas amostras.

- Além da acurácia, a aumentação de dados também melhorou a precisão e a sensibilidade

Tabela 5.3: Precisão para os experimentos com aumentação de dados.

Precisão									
Tamanho	Todas as Classes			COVID			Balanceado		
(%)	Normal	Pneumonia	COVID-19	Normal	Pneumonia	COVID-19	Normal	Pneumonia	COVID-19
10	0,9204	0,8504	**1.0000**	0,9309	0,8534	0,9231	0,8999	0,7274	0,9250
20	0,9090	0,8949	0,9773	0,9283	0,8824	0,9474	0,8292	0,8124	0,9000
30	0,9344	0,8972	0,8684	0,9339	0,8959	0,9512	0,8962	0,8301	0,8500
40	0,9425	0,8738	0,9483	0,9385	0,8919	0,9583	0,9151	0,8326	0,8974
50	0,9323	0,9067	0,9524	0,939	0,8712	0,9167	0,9027	0,8704	**0,9423**
60	0,9374	0,8961	0,9423	0,9457	0,8990	0,9194	0,9226	0,8854	0,8767
70	0,9344	0,9181	0,9701	0,9374	0,8989	0,9821	0,8992	0,9052	0,8846
80	0,9391	0,9361	0,9375	**0,9529**	0,9063	**0,9853**	0,9106	**0,9165**	0,8182
90	**0,9518**	0,9101	0,9595	0,9451	0,9307	0,9605	0,9211	0,8807	0,7717
100	0,9344	**0,9571**	0,9111	0,9344	**0,9571**	0,9111	**0,9232**	0,9037	0,7339

Tabela 5.4: Sensibilidade para os experimentos com aumentação de dados.

Sensibilidade									
Tamanho	Todas as Classes			COVID			Balanceado		
(%)	Normal	Pneumonia	COVID-19	Normal	Pneumonia	COVID-19	Normal	Pneumonia	COVID-19
10	0,9537	0,9091	0,2700	0,9593	0,9209	0,2400	0,8226	0,8939	0,3700
20	0,9706	0,8889	0,4300	0,9650	0,9226	0,3600	0,9164	0,7946	0,1800
30	0,9503	0,9108	0,6600	0,9740	0,9276	0,3900	0,9266	0,8721	0,3400
40	0,9446	0,9327	0,5500	0,9661	0,9310	0,4600	0,9254	**0,9040**	0,3500
50	0,9650	0,9158	0,6000	0,9571	0,9226	0,4400	0,9435	0,8822	0,4900
60	0,9650	0,9293	0,4900	0,9650	0,9293	0,5700	0,9424	0,8973	0,6400
70	0,9650	0,9242	0,6500	0,9638	0,9276	0,5500	**0,9571**	0,8519	0,6900
80	0,9763	0,9125	0,7500	0,9605	**0,9444**	0,6700	0,9548	0,8687	0,7200
90	0,9605	**0,9377**	0,7100	0,9729	0,9276	0,7300	0,9367	0,8704	0,7100
100	**0,9819**	0,9007	**0,8200**	**0,9819**	0,9007	**0,8200**	0,9379	0,8687	**0,8000**

dos classificadores.

- Entretanto, mesmo ao empregarmos amostras aumentadas, a sensibilidade da categoria *COVID-19* permaneceu extremamente baixa.

Dessa maneira, os experimentos realizados nos permitiram observar alguns dos benefícios e das limitações que as técnicas de aumentação de dados podem trazer para problemas de classificação. Isso nos direciona à investigação de novas metodologias que contribuam para uma melhor resolução dos problemas de classificação a serem tratados e à busca de métodos que permitam a superação das limitações que serão encontradas em cada abordagem. Essas serão tarefas realizadas ao longo deste trabalho.

Capítulo 6

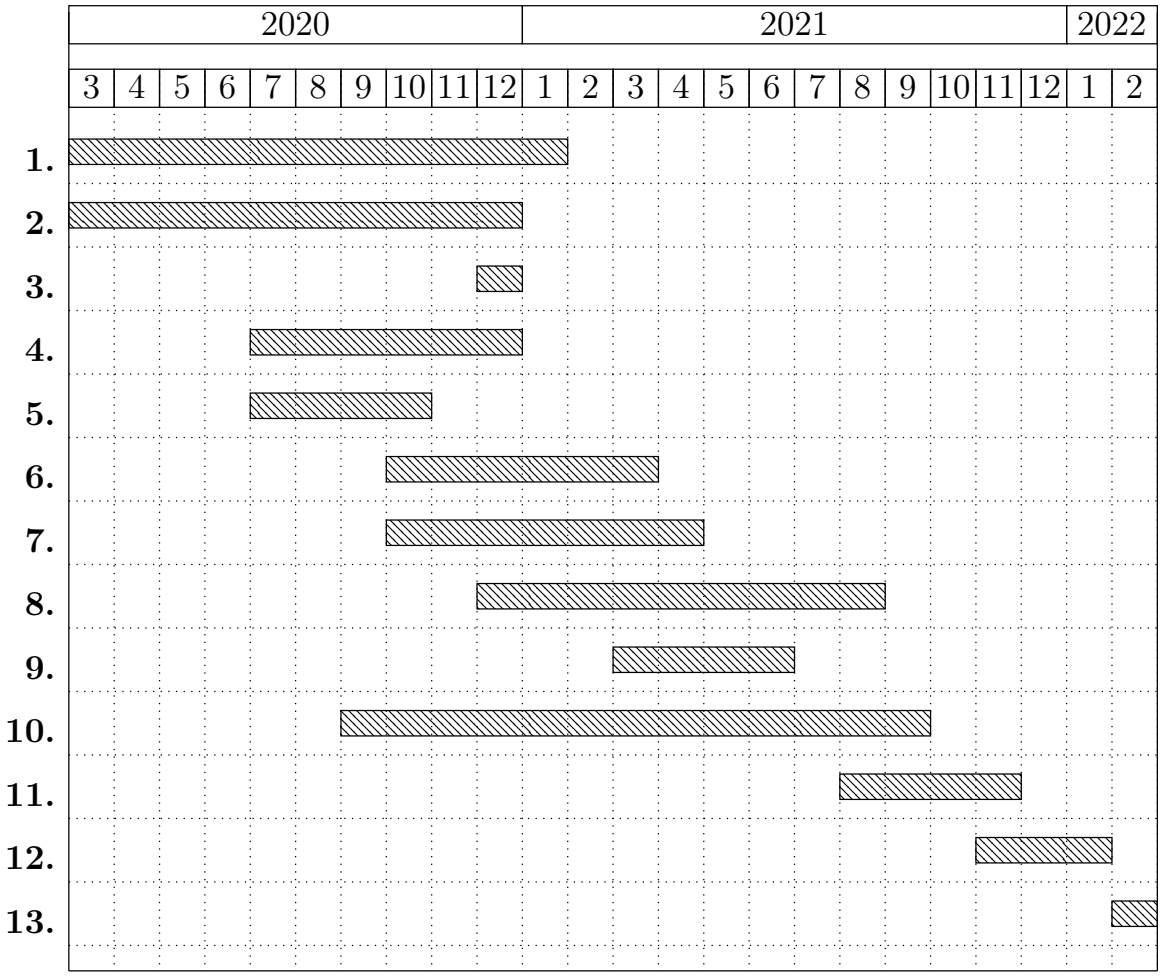
Plano de Trabalho e Cronograma de Execução

O plano de trabalho é composto pelas seguintes atividades:

1. Obtenção dos créditos obrigatórios em disciplinas do programa de mestrado.
2. Revisão bibliográfica em classificação de imagens e aumentação de dados.
3. Exame de Qualificação do Mestrado (EQM).
4. Pré-processamento das bases de dados.
5. Seleção de uma rede neural inicial.
6. Definição de arquiteturas para classificação de imagens para os casos de uso escolhidos.
7. Seleção de técnicas de aumentação de dados a serem empregadas e combinadas.
8. Aprimoramento das arquiteturas e das técnicas de aumentação escolhidas.
9. Participação do Programa de Estágio Docente (PED).
10. Realização de testes e análise dos resultados.
11. Documentação e publicação dos resultados.
12. Escrita da dissertação de mestrado.
13. Apresentação da dissertação de mestrado.

O cronograma de execução das atividades propostas, em um prazo de 24 meses, é apresentado na Tabela 6.1.

Tabela 6.1: Cronograma de atividades dividido em meses.



Bibliografia

[1] Dogs vs. Cats Redux: Kernels Edition. `https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition`, 2016.

[2] E. Andersson and R. Berglund. Evaluation of Data Augmentation of MR Images for Deep Learning, 2018. Retrieved from: `https://lup.lub.lu.se/student-papers/search/publication/8952747`. Accessed on: 03-November-2020.

[3] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein GAN. *arXiv:1701.07875*, 2017.

[4] A. Atapour-Abarghouei and T. P. Breckon. Real-Time Monocular Depth Estimation Using Synthetic Data with Domain Adaptation via Image Style Transfer. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2800–2810, 2018.

[5] J. G. A. Barbedo. Plant disease identification from individual lesions and spots using deep learning. *Biosystems Engineering*, 180:96–107, 2019.

[6] J. G. A. Barbedo, L. V. Koenigkan, B. A. Halfeld-Vieira, R. V. Costa, K. L. Nechet, C. V. Godoy, M. L. Junior, F. R. A. Patricio, V. Talamini, L. G. Chitarra, S. A. S. Oliveira, A. K. N. Ishida, J. M. C. Fernandes, T. T. Santos, F. R. Cavalcanti, D. Terao, and F. Angelotti. Annotated Plant Pathology Databases for Image-Based Detection and Recognition of Diseases. *IEEE Latin America Transactions*, 16(6):1749–1757, 2018.

[7] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. VGGFace2: A dataset for recognising faces across pose and age. In *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pages 67–74, 2018.

[8] A. Chung. Actualmed COVID-19 Chest X-ray Dataset Initiative. `https://github.com/agchung/Actualmed-COVID-chestxray-dataset`, 2020.

[9] A. Coates, A. Ng, and H. Lee. An Analysis of Single-Layer Networks in Unsupervised Feature Learning. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 215–223, 2011.

[10] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC). In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 168–172, 2018.

[11] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi. COVID-19 Image Data Collection: Prospective Predictions Are the Future. *arXiv:2006.11988*, 2020.

[12] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le. AutoAugment: Learning Augmentation Strategies from Data. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 113–123, 2019.

[13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009.

[14] T. DeVries and G. W. Taylor. Dataset Augmentation in Feature Space. *arXiv:1702.05538*, 2017.

[15] T. DeVries and G. W. Taylor. Improved Regularization of Convolutional Neural Networks with Cutout. *arXiv:1708.04552*, 2017.

[16] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon. Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 2106–2112, 2011.

[17] Y. Em, F. Gag, Y. Lou, S. Wang, T. Huang, and L.-Y. Duan. Incorporating Intra-Class Variance to Fine-Grained Visual Recognition. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1452–1457, 2017.

[18] L. Engstrom. Fast style transfer. `https://github.com/lengstrom/fast-style-transfer/`, 2016.

[19] D. Erhan, Y. Bengio, A. Courville, P.-A. Manzagol, P. Vincent, and S. Bengio. Why Does Unsupervised Pre-training Help Deep Learning? *Journal of Machine Learning Research*, 11(Feb):625–660, 2010.

[20] L. Fei-Fei, R. Fergus, and P. Perona. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):594–611, 2006.

[21] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan. GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification. *Neurocomputing*, 321:321–331, 2018.

[22] X. Gastaldi. Shake-Shake regularization. *arXiv:1705.07485*, 2017.

[23] M. Geng, K. Xu, B. Ding, H. Wang, and L. Zhang. Learning data augmentation policies using augmented random search. *arXiv:1811.04768*, 2018.

[24] G. Ghiasi, H. Lee, M. Kudlur, V. Dumoulin, and J. Shlens. Exploring the structure of a real-time, arbitrary neural artistic stylization network. *arXiv:1705.06830*, 2017.

[25] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio. Challenges in Representation Learning: A report on three machine learning contests. In *International Conference on Neural Information Processing (ICONIP)*, pages 117–124, 2013.

[26] G. Griffin, A. Holub, and P. Perona. Caltech-256 Object Category Dataset. 2007.

[27] R. Gupta, R. Hosfelt, S. Sajeev, N. Patel, B. Goodman, J. Doshi, E. Heim, H. Choset, and M. Gaston. xBD: A Dataset for Assessing Building Damage from Satellite Imagery. *arXiv:1911.09296*, 2019.

[28] S. Gururangan, A. Marasović, S. Swayamdipta, K. Lo, I. Beltagy, D. Downey, and N. Smith. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 8342–8360, 2020.

[29] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[30] D. Hendrycks and T. G. Dietterich. Benchmarking Neural Network Robustness to Common Corruptions and Surface Variations. *arXiv:1807.01697*, 2018.

[31] J. Hu, L. Shen, and G. Sun. Squeeze-and-Excitation Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7132–7141, 2018.

[32] G. Huang, Z. Liu, G. Pleiss, L. Van Der Maaten, and K. Weinberger. Convolutional Networks with Dense Connectivity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[33] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely Connected Convolutional Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017.

[34] H. Inoue. Data Augmentation by Pairing Samples for Images Classification. *arXiv:1801.02929*, 2018.

[35] S. Ioffe and C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167*, 2015.

[36] P. T. Jackson, A. A. Abarghouei, S. Bonner, T. P. Breckon, and B. Obara. Style Augmentation: Data Augmentation via Style Randomization. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, pages 83–92, 2019.

[37] A. Kamilaris and F. Prenafeta-Boldú. A review of the use of convolutional neural networks in agriculture. *The Journal of Agricultural Science*, 156(3):312–322, 2018.

[38] G. Koch. Siamese Neural Networks for One-Shot Image Recognition. Master's thesis, Graduate Department of Computer Science University of Toronto, 2015.

[39] J. Krause, J. Deng, M. Stark, and L. Fei-Fei. Collecting a Large-Scale Dataset of Fine-Grained Cars. 2013.

[40] N. C. Krishnan and D. J. Cook. Activity Recognition on Streaming Sensor Data. *Pervasive and Mobile Computing*, 10:138–154, 2014.

[41] A. Krizhevsky and G. Hinton. Learning Multiple Layers of Features from Tiny Images. Technical report, University of Toronto, 2009.

[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Neural Information Processing Systems (NIPS)*, pages 1097–1105, 2012.

[43] Y. LeCun. The MNIST database of handwritten digits. `http://yann.lecun.com/exdb/mnist/`, 1998.

[44] R. S. Lee, F. Gimenez, A. Hoogi, and D. Rubin. Curated Breast Imaging Subset of DDSM. *The Cancer Imaging Archive*, 8, 2016.

[45] J. Lemley, S. Bazrafkan, and P. Corcoran. Smart Augmentation - Learning an Optimal Data Augmentation Strategy. *IEEE Access*, 5:5858–5869, 2017.

[46] X. Li and X. Wu. Constructing Long Short-Term Memory based Deep Recurrent Neural Networks for Large Vocabulary Speech Recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4520–4524, 2015.

[47] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal Loss for Dense Object Detection. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.

[48] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision (ECCV)*, pages 740–755, 2014.

[49] W. Liu, Y. Wen, Z. Yu, and M. Yang. Large-Margin Softmax Loss for Convolutional Neural Networks. In *International Conference on Machine Learning (ICML)*, volume 48, page 507–516, 2016.

[50] R. G. Lopes, D. Yin, B. Poole, J. Gilmer, and E. D. Cubuk. Improving Robustness Without Sacrificing Accuracy with Patch Gaussian Augmentation. *arXiv:1906.02611*, 2019.

[51] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek. The Japanese Female Facial Expression (JAFFE) Database. In *International Conference on Automatic Face and Gesture Recognition (FG)*, pages 14–16, 1998.

[52] S. Maji, E. Rahtu, J. Kannala, M. Blaschko, and A. Vedaldi. Fine-Grained Visual Classification of Aircraft. *arXiv:1306.5151*, 2013.

[53] H. Mania, A. Guy, and B. Recht. Simple random search provides a competitive approach to reinforcement learning. *arXiv:1803.07055*, 2018.

[54] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley. Least Squares Generative Adversarial Networks. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2794–2802, 2017.

[55] A. Mikołajczyk and M. Grochowski. Data augmentation for improving deep learning in image classification problem. In *International Interdisciplinary PhD Workshop (IIPhDW)*, pages 117–122, 2018.

[56] T. N. Minh, M. Sinn, H. T. Lam, and M. Wistuba. Automated Image Data Preprocessing with Deep Reinforcement Learning. *arXiv:1806.05886*, 2018.

[57] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng. Reading Digits in Natural Images with Unsupervised Feature Learning. 2011.

[58] K. Nichol. Kaggle Dataset: Painter by Numbers. `https://www.kaggle.com/c/painter-by-numbers`, 2016.

[59] M.-E. Nilsback and A. Zisserman. Automated Flower Classification over a Large Number of Classes. In *Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, pages 722–729, 2008.

[60] A. Odena, C. Olah, and J. Shlens. Conditional Image Synthesis with Auxiliary Classifier GANs. In *International Conference on Machine Learning (ICML)*, pages 2642–2651, 2017.

[61] F. Perez, C. Vasconcelos, S. Avila, and E. Valle. Data Augmentation for Skin Lesion Analysis. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*, pages 303–311. Springer, 2018.

[62] L. Perez and J. Wang. The Effectiveness of Data Augmentation in Image Classification Using Deep Learning. *arXiv:1712.04621*, 2017.

[63] A. Radford, L. Metz, and S. Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv:1511.06434*, 2015.

[64] Radiological Society of North America. COVID-19 Radiography Database. `https://www.kaggle.com/tawsifurrahman/covid19-radiography-database`, 2019.

[65] Radiological Society of North America. RSNA Pneumonia Detection Challenge. `https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data`, 2019.

[66] J. Rajendran, A. Irpan, and E. Jang. Meta-Learning Requires Meta-Augmentation. *arXiv:2007.05549*, 2020.

[67] A. J. Ratner, H. Ehrenberg, Z. Hussain, J. Dunnmon, and C. Ré. Learning to Compose Domain-Specific Transformations for Data Augmentation. In *Advances in Neural Information Processing Systems (NIPS)*, pages 3236–3246, 2017.

[68] E. Real, A. Aggarwal, Y. Huang, and Q. V. Le. Regularized Evolution for Image Classifier Architecture Search. In *Conference on Artificial Intelligence (AAAI)*, volume 33, pages 4780–4789, 2019.

[69] A. Ruano Miralles. An open-source development environment for Self-driving vehicles. Master's thesis, Universitat Oberta de Catalunya, 2017.

[70] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting Visual Category Models to New Domains. In *European Conference on Computer Vision (ECCV)*, pages 213–226, 2010.

[71] M. Safdar, S. AlKobaisi, and F. Zahra. A Comparative Analysis of Data Augmentation Approaches for Magnetic Resonance Imaging (MRI) Scan Images of Brain Tumor. *Acta Informatica Medica*, 28:29–36, 2020.

[72] C. Shorten and T. M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1):60, 2019.

[73] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556*, 2014.

[74] K. Sohn, D. Berthelot, C.-L. Li, Z. Zhang, N. Carlini, E. D. Cubuk, A. Kurakin, H. Zhang, and C. Raffel. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. *arXiv:2001.07685*, 2020.

[75] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[76] C. Summers and M. J. Dinneen. Improved Mixed-Example Data Augmentation. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1262–1270, 2019.

[77] Q. Sun, Y. Liu, T.-S. Chua, and B. Schiele. Meta-Transfer Learning for Few-Shot Learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[78] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In *Conference on Artificial Intelligence (AAAI)*, page 4278–4284, 2017.

[79] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the Inception Architecture for Computer Vision. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

[80] R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer Science & Business Media, 2010.

[81] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding. Embracing Imperfect Datasets: A Review of Deep Learning Solutions for Medical Image Segmentation. *Medical Image Analysis*, page 101693, 2020.

[82] J. Uhrig, N. Schneider, L. Schneider, U. Franke, T. Brox, and A. Geiger. Sparsity Invariant CNNs. In *International Conference on 3D Vision (3DV)*, pages 11–20, 2017.

[83] L. Wang and A. Wong. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images. *arXiv:2003.09871*, 2020.

[84] Y. Wang, X. Pan, S. Song, H. Zhang, G. Huang, and C. Wu. Implicit Semantic Data Augmentation for Deep Networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 12635–12644, 2019.

[85] Z. Wang, T. Schaul, M. Hessel, H. Van Hasselt, M. Lanctot, and N. De Freitas. Dueling Network Architectures for Deep Reinforcement Learning. In *International Conference on Machine Learning (ICML)*, page 1995–2003, 2016.

[86] Y. Wen, K. Zhang, Z. Li, and Y. Qiao. A Discriminative Feature Learning Approach for Deep Face Recognition. In *European Conference on Computer Vision (ECCV)*, pages 499–515, 2016.

[87] Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata. Zero-Shot Learning - A Comprehensive Evaluation of the Good, the Bad and the Ugly. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(9):2251–2265, 2019.

[88] L. Xie, J. Wang, Z. Wei, M. Wang, and Q. Tian. DisturbLabel: Regularizing CNN on the Loss Layer. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4753–4762, 2016.

[89] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He. Aggregated Residual Transformations for Deep Neural Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1492–1500, 2017.

[90] Y. Yamada, M. Iwamura, T. Akiba, and K. Kise. ShakeDrop Regularization for Deep Residual Learning. *IEEE Access*, 7:186126–186136, 2019.

[91] S. Zagoruyko and N. Komodakis. Wide Residual Networks. In *British Machine Vision Conference (BMVC)*, pages 87.1–87.12. BMVA Press, 2016.

[92] X. Zhang, Z. Wang, D. Liu, and Q. Ling. DADA: Deep Adversarial Data Augmentation for Extremely Low Data Regime Classification. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2807–2811, 2019.

[93] Z. Zhang and M. Sabuncu. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. In *Advances in Neural Information Processing Systems (NIPS)*, pages 8778–8788, 2018.

[94] X. Zheng, T. Chalasani, K. Ghosal, S. Lutz, and A. Smolic. STaDA: Style Transfer as Data Augmentation. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP)*, 2019.

[95] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang. Random Erasing Data Augmentation. *Conference on Artificial Intelligence (AAAI)*, 34, 2017.

[96] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. Places: A 10 million Image Database for Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(6):1452–1464, 2017.

[97] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2223–2232, 2017.

[98] X. Zhu, Y. Liu, J. Li, T. Wan, and Z. Qin. Emotion Classification with Data Augmentation Using Generative Adversarial Networks. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pages 349–360, 2018.

[99] B. Zoph and Q. V. Le. Neural Architecture Search with Reinforcement Learning. *arXiv:1611.01578*, 2017.

[100] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le. Learning Transferable Architectures for Scalable Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8697–8710, 2018.