

Desenvolvendo um Ensemble de Redes Profundas para identificação de Lentes Gravitacionais: Aplicação em Regime de Poucos dados

Deep Ensemble to search for Strong Gravitational Lensing: application to small training sets

Icaro F. M. de Castro^{1,2,*}, Kayque Teles², Clécio R. Bom^{2,3}, e Tatiana Escovedo¹
¹ Departamento de Informática Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Brasil

² Centro Brasileiro de Pesquisas Físicas
Rua Dr. Xavier Sigaud, 150, Ed. César Lattes,
Urca, Rio de Janeiro, RJ. CEP: 22290-180, Brasil e

³ Centro Federal de Educação Tecnológica Celso Suckow da Fonseca,
Rodovia Mário Covas, lote J2, quadra J
Distrito Industrial de Itaguaí,

Itaguaí – RJ. CEP: 23810-000, Brasil

Submetido: 28/02/2021 Aceito: 16/06/2021

Resumo: O treinamento de redes neurais profundas com pequenos conjuntos de dados é um desafio em diversos casos. Isso pode ocorrer quando o fenômeno de interesse é imprevisível ou extremamente raro. Neste trabalho, analisamos o desempenho da classificação realizada pelas redes profundas ResNet 50, EfficientNet B2 e do Ensemble entre elas. O conjunto de dados utilizado é um banco de imagens simuladas de um fenômeno físico bastante raro: o lenteamento gravitacional. Foram feitos ajustes nos modelos, com o objetivo de obter o melhor desempenho com o menor conjunto de dados de treinamento. Testamos os modelos com diferentes quantidades de dados, analisando o desempenho de cada um a partir da métrica AUC (área sob a curva ROC). O melhor desempenho em um conjunto restrito a 80 imagens de treino foi obtido através de um ensemble dos dois modelos de redes neurais profundas com uma AUC 0,796.

Palavras-chave: ResNet, EfficientNet, Ensemble, Lente gravitacional.

Abstract: The training of deep neural networks with small data sets is a challenge in many cases, particularly when a given phenomenon is extremely rare. This paper analyses the performance of the deep networks ResNet50, EfficientNet B2 and their Ensemble. The dataset used is a image bank of simulated images over a rare physical system: Strong Gravitational Lensing. The models were adjusted in order to obtain the best performance with the smallest amount of training data. We made tests with different amounts of data, analyzing the performance of each model based on the AUC metric (area under ROC curve). The best performance using a training set of only 80 training images was obtained from an Ensemble of the two models of deep neural networks with an AUC of 0.796.

Keywords: ResNet, EfficientNet, Ensemble, Gravitational Lens.

1. INTRODUÇÃO

O fenômeno de lenteamento gravitacional consiste na deflexão da luz de uma fonte luminosa por um corpo massivo [1–4]. O campo gravitacional do objeto massivo atua como uma lente que distorce a imagem do corpo longínquo. O fenômeno é uma consequência da curvatura do espaço tempo, prevista pela teoria de gravitação de Einstein [5–7]. No entanto, foi somente em 1979 que o fenômeno foi verificado numa imagem dupla do quasar QSO 0957+561 [8]. Anteriormente, o objeto tinha o seu espectro estudado como o de uma dupla de quasares gêmeos [9].

No presente momento, são conhecidos da ordem de centenas de sistemas de lenteamento forte confirmados e alguns milhares de candidatos. Com o advento do *Vera Rubin Ob-*

servatory, [10–12], espera-se observar da ordem de 10^5 sistemas de lentes em escala galáctica [13], dos quais cerca de 120 deverão ser sistemas de múltiplos planos. E consequentemente, poderá fomentar da pesquisa na área de Lentes Gravitacionais em dois planos (*Double Source Plane Lensing*, DSPL, e.g. [14, 15]). Este fenômeno tem o potencial de ser complementar a cosmologia com *time-delay*, podendo melhorar a figura de mérito da Energia Escura em cerca de 40%, considerando uma amostra menor que 100 sistemas com baixos *redshifts* [14]. Com a análise e modelagem detalhada das subestruturas dos sistemas na escala de galáxias, também será possível obter limites de modelos de matéria escura fria e matéria escura morna [16], testando assim o atual paradigma da Cosmologia.

O uso de técnicas de aprendizagem profunda em astrofísica vem sendo largamente difundido ao longo dos últimos anos (veja ,e.g., [17–23]), principalmente para tarefas de classificação, em um caso típico de grande volume de dados (Big Data), assim como em levantamentos de grande

*Electronic address: icaro-fmc@hotmail.com

área. Dado o volume de dados, sua variedade e, confiabilidade, a astrofísica se apresenta como um laboratório para este tipo de técnica, na busca por lentes gravitacionais [24–26]. No entanto, como lentes gravitacionais são objetos raros, a busca por estes objetos é dominada por falsos positivos. Em geral, para desenvolver tais classificadores, uma escolha comum é elaborar simulações para treinar algoritmos.

Nas últimas décadas, um tipo de rede neural profunda, conhecida como rede neural convolucional, emergiu como uma solução de maior acurácia para problemas de classificação e visão computacional [27–30]. Esta técnica se inspira na forma como mamíferos processam informação visual para produzir modelos matemáticos, utilizando milhares ou milhões de parâmetros, definidos após um processo de treinamento com dados, que simulam as conexões neuronais [31].

As redes neurais convolucionais são aquelas que se utilizam de um kernel ou filtro que é convolucionado com os dados processados através da rede, buscando identificar características nas imagens, que possibilitem diversas análises tais como classificação e regressão[20].

Nesta contribuição, utilizaremos as redes neurais convolucionais *ResNet 50*[22], *EfficientNet B2*[32] e construiremos um *Ensemble* destes dois modelos a fim de realizar a classificação de objetos astrofísicos conhecidos como lentes gravitacionais, a partir de imagens astronômicas em formato fits. O foco deste trabalho é analisar o regime de poucos dados, com as seguintes motivação: i) avaliar a generalização da rede. À medida em que as redes aprendem com uma menor quantidade de dados isso aumenta seu potencial de adaptação em diferentes datasets, evitando a necessidade de construção de grandes datasets de treino/teste, com grande variedade e controle de qualidade; ii) Aplicabilidade em base de dados astronômicos que não cubram uma grande área muito grande onde sistemas de lentes gravitacionais conhecidos sejam raros até mesmo para validar o código. iii) a possibilidade de elaborar uma metodologia, que possa ser empregada na construção de modelos de busca de lentes gravitacionais em poucos dados e portanto independentes da construção de um grande dataset simulado para treinamento e assim, diretamente aplicável a dados reais.

Para a construção da metodologia proposta, utilizaremos um dataset de simulações, já consolidado na literatura, originalmente criado para o primeiro *Gravitational Lens Finding Challenge*[24]. O interesse na escolha deste dataset simulado se deve a sua grande variedade e efeitos observacionais considerados. Uma vez que nosso objetivo é avaliar o limite de poucos dados, é necessário partir de um dataset suficientemente grande e variado.

Esse trabalho está organizado da seguinte maneira: na seção 2 é feita uma breve discussão teórica acerca do efeito forte de lenteamento gravitacional. Em seguida, na seção 3, fazemos um breve resumo das técnicas atuais utilizadas e do estado-da-arte da busca por lentes gravitacionais gravitacionais. Na seção 4, metodologia, discutimos as motivações para o uso do *deep learning* como estratégia para realizar a classificação de imagens, explicando os principais aspectos das redes neurais convolucionais e as métricas utilizadas para a avaliação dos resultados. Por fim, é feita uma análise técnica do conjunto de dados utilizado, na seção 5 seguida

da apresentação dos resultados na seção 6 e discussões finais na seção 7.

2. EFEITO FORTE DE LENTES GRAVITACIONAIS

Segundo a teoria da Relatividade Geral [6, 7], a luz é defletida na presença de um campo gravitacional não-homogêneo [33, 34]. O objeto massivo que gera esse campo gravitacional é conhecido como "lente" e o objeto que teve a sua luz defletida é chamado de "fonte". A imagem distorcida da fonte devido a deflexão da luz se conhece por efeito de lenteamento gravitacional.

Consideremos o caso em que o objeto M , que serve como lente, é pequeno o bastante em comparação à distância viajada pela luz emitida pela fonte S e que também não é extenso de modo que possa ser considerado uma lente pontual. Situando o objeto M ao longo da linha de visada, é descrito um ângulo β com a linha que liga a fonte S diretamente ao observador. No entanto, devido ao campo gravitacional nas proximidades da massa M , os raios de luz emitidos pela fonte S sofrem uma deflexão α . Esta deflexão, faz a imagem I da fonte S ser percebida pelo observador na posição angular θ , conforme figura 1, de escalas angulares exageradas propositalmente.

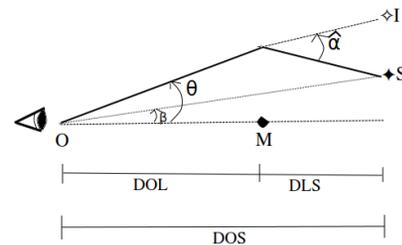


Figura 1: Representação esquemática do caminho descrito por um raio de luz emitido por uma fonte S , na presença de uma grande massa M . Se não houvesse a massa M , a fonte S seria vista com o ângulo β em relação à linha de visada.

Seja D_{OL} a distância de diâmetro angular [35] do observador à lente, D_{LS} a distância de diâmetro angular da lente à fonte e D_{OS} a referida distância do observador à fonte, chegamos a equação 1, partindo das relações angulares apresentadas na figura 1.

$$\theta D_{OS} = \beta D_{OS} + \hat{\alpha} D_{LS}. \quad (1)$$

Introduzindo o ângulo de deflexão como $\alpha = \hat{\alpha} \frac{D_{LS}}{D_{OS}}$ [4], obtemos a equação da lente

$$\beta = \theta - \alpha, \quad (2)$$

que descreve a relação entre as posições reais e aparentes de uma imagem. Tipicamente, quanto à magnitude da distorção da imagem, o efeito de lente divide-se em dois regimes:

- **Lenteamento Fraco:** Considera-se que o lenteamento é fraco quando, na aproximação em que o tamanho do objeto que serve de lente é pequeno em comparação

à trajetória da luz, os ângulos de deflexão α são pequenos e é relativamente fácil discernir as posições verdadeiras dos objetos lenteados. É um efeito apenas medido estatisticamente, considerando um grande conjunto de fontes que apresentam algum tipo de padrão espacial [36].

- **Lenteamento Forte:** Por sua vez, considera-se que o efeito de lente é forte quando as deflexões ocorrem em grandes ângulos α , o que resulta em grandes distorções das imagens [36], propiciando o aparecimento de imagens múltiplas do mesmo objeto. Além disso, as imagens do objeto lenteado podem ser magnificadas [37], do mesmo modo que uma lente convexa pode concentrar a luz de um objeto distante, num ponto em que o observador verá o objeto mais brilhante do que realmente é [34].

Esse efeito demonstrou-se uma ferramenta importante para estudar a estrutura interna de aglomerados e a distribuição de matéria do universo [4, 36].

Como já exposto, as distorções de imagens devido ao efeito forte de lente se manifestam na forma de imagens múltiplas do objeto longínquo. No entanto, conforme o ângulo descrito pela linha que liga o observador à lente e a linha que liga a lente e a fonte diminui, as múltiplas imagens vão se aproximando e adquirindo a forma de arcos ao redor da lente. Em particular, quando esse ângulo tende à zero, isto é, há um alinhamento perfeito entre observador, lente e fonte, os arcos se fundem, formando uma estrutura anelar (ver figura 2) conhecida como Anel de Einstein [1].

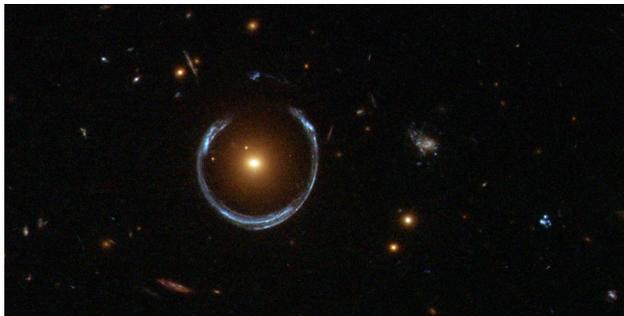


Figura 2: Imagem lenteada (galáxia azul) por uma galáxia vermelha luminosa (LRG), formando a estrutura do Anel de Einstein. Créditos: ESA/Hubble - NASA.

Isso se deve ao fato de que, no trajeto da luz, pode haver mais de uma trajetória possível conectando a fonte com o observador, de modo que vários raios de luz cheguem a este [33]. Matematicamente, pode-se considerar que, no caso de uma lente pontual, o ângulo de deflexão seja representado conforme a equação [4, 38]:

$$\alpha = \frac{D_{LS}}{D_{OS}D_{OL}} \frac{4GM}{c^2\theta}, \quad (3)$$

onde G é a constante da Gravitação Universal, M a massa da lente e c a velocidade da luz no vácuo.

Usando este resultado na equação 2, podemos reescreve-la como $\beta = \theta - \frac{\theta_E^2}{\theta}$, em que a quantidade θ_E é conhecida como

Ângulo de Einstein. Pode-se verificar que essa equação possui duas soluções à priori, o que indica o efeito de formação de imagens duplas devido ao lenteamento gravitacional. No caso em que $\beta = 0$, temos devidamente a sobreposição das duas soluções, em que o ângulo de Einstein indica o raio do Anel de Einstein, definido conforme a equação 4

$$\theta_E = \sqrt{\frac{D_{LS}}{D_{OS}D_{OL}} \frac{4GM}{c^2}}. \quad (4)$$

A presença de um anel de Einstein permite inferir diversas características da lente, como a quantidade de massa da lente situada dentro do raio de Einstein [39]. A curvatura deste anel ou, a formação dos chamados arcos gravitacionais, dependendo das posições da lente/fonte/observador, motivou métodos de medidas de lentes [40–42], que eventualmente foram propostas como estimadores para busca desses sistemas [25, 43, 44].

O efeito forte de lente também possui diversas aplicações em Cosmologia e Astrofísica, como a inferência da constante de Hubble, ao medir as diferenças de tempo entre pares de imagens múltiplas [39, 45] e a distribuição de matéria em aglomerados [46, 47].

Em outros regimes, o efeito de lente também pode ser usado para limitar o tamanho de componentes de quasares e a estrutura de seus discos de acreção [48]; servir como telescópios naturais para observar objetos que seriam, de outra forma, indetectáveis [e.g., 49–51] estudar a distribuição de matéria escura nos objetos que servem de lente [45]; encontrar exoplanetas [52]; ou, até mesmo, a possibilidade do lenteamento pela matéria escura [53].

3. BUSCA DE LENTES GRAVITACIONAIS

Devido a grande quantidade de aplicações e o grande interesse da comunidade científica por sistemas de lentes gravitacionais diversos esforços foram realizados a fim de se buscar esses sistemas, em particular em levantamentos de grande área de imagens, os chamados *surveys*.

Uma das iniciativas pioneiras em obtenção de novas imagens do espaço, para uso na astrofísica e cosmologia, foi o *Sloan Digital Sky Survey* (SDSS) [54], sendo utilizado por diversos grupos para encontrar candidatos de lentes [43, 55–57].

Outros *surveys* foram utilizados para encontrar efeito forte de lente gravitacional, incluindo análises na banda de rádio [58, e.g.]. Outra base de dados explorada para este fim, foi o *Kilo-Degree Survey* (KiDS) que inclui em seus dados imagens do telescópio VISTA, localizado no hemisfério sul [59, 60]. Cabe mencionar, ainda, o *Dark Energy Survey* (DES), uma iniciativa que realizou um levantamento sistemático de 5.000 graus quadrados, a fim de mapear o céu e conduzir diversas análises, incluindo o efeito de lenteamento fraco [61]. Diversas buscas foram realizadas, a fim de encontrar sistemas de lentes gravitacionais fortes [62–64].

Alguns dos problemas na busca deste tipo de fenômeno se dão a medida que os objetos, como galáxias, se tornam cada vez mais difíceis de distinguir em altos desvios para o vermelho, i.e. mais distantes, conforme se tornam

menos brilhantes e com menor tamanho angular, dificultando a inspeção visual [65, 66]. Além disso conforme o volume de dados disponíveis cresce esse tipo de abordagem se torna ineficiente para análise massiva de dados. Uma alternativa a inspeção visual é o desenvolvimento de métodos computacionais de busca de sistemas de lentes automatizadas por Aprendizado de Máquina (ou *Machine Learning*) e *Deep Learning* [24, 25, 43, 67–70], sendo este último um ramo do *Machine Learning* que dispõe de redes com um grande número de camadas e parâmetros, tipicamente utilizando requerendo uso intensivo e computação e GPUs e [71]. Métodos de *Deep Learning*, e especificamente um subtipo, as Redes Neurais Convolucionais (da sigla em inglês CNNs) são consideradas o estado-da-arte do reconhecimento de padrões em imagens e, portanto particularmente interessantes na busca de efeito forte de lente em imagens de galáxias. As camadas convolucionais de uma rede neural profunda podem reduzir uma imagem complexa à representações bem simples, realizando a detecção de suas características com mais eficiência e permitindo o aprendizado sobre todo o conjunto de dados [72].

O uso de CNNs em dados astronômicos tem apresentado diversas aplicações. Em um trabalho pioneiro Hála (2014) [73] desenvolveu classificação espectral de galáxias, quasares e estrelas. CNNs também foram aplicadas para a classificação morfológica de galáxias [74–76] e estimação fotométrica de redshifts [77]. Nos dados do KiDS, foi realizada uma busca de efeito forte de lente de forma automática utilizando *deep learning* [60, 78, 79]. Nos dados do DES, também já foi encontrado efeito forte de lente em altos desvios pro vermelho, utilizando a busca automatizada das redes neurais convolucionais [63].

Devido ao crescente interesse da astrofísica pela análise automatizada em imagens e catálogos de *surveys*, surgiram diversas competições de ciência de dados no intuito de estimular o desenvolvimento de códigos robustos. Entre essas competições está o *Galaxy Challenge*, um desafio hospedado na plataforma *Kaggle*, que propôs o desenvolvimento de um código capaz de automatizar a classificação de galáxias de um *dataset*, auxiliado pelo projeto *Galaxy Zoo* [80]. Este projeto se dispõe a fazer um catálogo morfológico de galáxias a partir de dados do SDSS e do *Dark Energy Camera Legacy Survey* (DECaLS)[81]. A submissão vencedora, por Sander Dieleman, utilizou um *ensemble* de 17 CNNs, sendo a maior delas uma rede com 42 milhões de parâmetros.

Uma competição mais recente, focada em efeito forte de lente, é o *The Strong Gravitational Lens Finding Challenge*, que propôs aos participantes a elaboração de um código capaz de realizar a classificação de imagens de lentes gravitacionais [24]. Para isto, na primeira edição foi oferecido um *dataset* próprio aos competidores, com imagens simuladas de galáxias, onde poderia haver, ou não, efeito de lenteamento. Parte dessas imagens simula a situação *space-based*, apresentando imagens como as retiradas de telescópios espaciais. A outra parte simula a situação *ground-based*, isto é, com imagens que reproduzem condições observacionais de telescópios localizados na Terra.

A CNN que ganhou a primeira edição dessa competição foi a CMU *Deep-Lens* [26], um código baseado nas redes neurais residuais (*ResNet*, ver seção 4.1.1), que possuem fa-

cidade para processar imagens altamente ruidosas [22]. Já a segunda edição contou com mais imagens simuladas, dessa vez aproximando-se das observações do Euclid Survey [82], e solicitou dos competidores, não apenas a identificação das lentes gravitacionais, como também uma estimação para o raio de Einstein das imagens observadas. A CNN vencedora foi a *Cast Efficient AUC*, por de Bom et. al (in prep.).

4. METODOLOGIA

Ao longo da década passada, as redes neurais convolucionais obtiveram grande êxito no reconhecimento de imagens e vídeos. Esse advento se tornou possível graças ao aumento do poder computacional, proporcionado pelas GPUs e pelos processamentos em *clusters* (sistemas distribuídos em grande escala) [83]. Outro fator relevante foi a disponibilização de grandes repositórios públicos de imagens como o *ImageNet* e os *surveys* de imagens astronômicas. O *ImageNet* é uma das maiores bases de dados já criada para uso público, consistindo em um conjunto de dezenas de milhões de imagens de objetos diversos, que vão desde mamíferos à ferramentas [84].

Um desafio importante para a evolução das arquiteturas profundas de reconhecimento de imagens, utilizando o referido *dataset*, foi o *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) [85, 86]. Esse desafio foi uma das primeiras junções entre a disponibilização pública de uma base de dados e a realização de uma competição recorrente, e sistemática, para visão computacional. O *dataset* desse desafio consiste em um conjunto de milhões de imagens organizadas conforme uma hierarquia *WordNet* [87], que promove o desenvolvimento e comparação de algoritmos de identificação de objetos. O fato da competição ter tido esse caráter recorrente permite acompanhar e discutir o progresso das inovações apresentadas em cada uma de suas edições [85].

Na edição de 2014 do ILSVRC, a equipe que obteve a primeira e a segunda colocação desenvolveu o modelo de CNN conhecido como VGG [86]. Esse modelo foi desenvolvido visando identificar o impacto do aumento da profundidade da rede neural na acurácia obtida no reconhecimento de imagens em grande escala. Foram realizados diversos testes com redes de profundidades e configurações diferentes [86].

Nos diversos desafios de reconhecimento de imagem, comumente, os *datasets* utilizados consistem de dezenas ou centenas de milhares de imagens dos objetos a serem classificados. As CNNs, frequentemente, dependem desta ordem de magnitude de dados para realizar a classificação com algum potencial real de predição [88]. Por outro lado, a inspeção visual de humanos tende a ser altamente efetiva, mesmo quando há uma quantidade mínima de dados. Ao se defrontar com um novo objeto, o processo de inspeção visual consegue definir uma nova classe para este, ou situá-lo numa classe já previamente estabelecida [89]. Já as redes neurais artificiais convolucionais profundas tradicionais tem performance similar a uma previsão aleatória a partir de apenas um único dado [90].

No intuito de analisar o problema da classificação com poucos dados, surgiu o ramo da inteligência artificial conhecido por *few-shot learning*. A partir de um conhecimento previamente adquirido, um método de *few-shot learning* pode rapidamente generalizar informações partindo de poucos exemplos, que dispõem de uma tabela-verdade. Essa capacidade é de interesse em diversas aplicações. Por exemplo, para uso na robótica, através da detecção visual de eventos e modelagem de linguagem [91, 92]. Os algoritmos de *few-shot-learning* minimizam a necessidade de se coletar grandes quantidades de dados supervisionados de um fenômeno ou objeto, possibilitando a realização de previsões precisas acerca destes [89, 93]. Desta forma, torna-se viável utilizar as redes neurais para estudar fenômenos que possuem poucos exemplos conhecidos, como os sistemas reais e confirmados de efeito forte de lente gravitacional sem necessidade de grandes simulações. Neste trabalho, no entanto, não adotamos um algoritmo de *few-shot learning*. Analisamos o limite de redes tradicionais e o uso de um *Ensemble*, i.e. uma combinação de redes, como estratégia para analisar a fronteira e os limites de poucos de dados nestas redes em que se é possível realizar previsões precisas.

4.1. Modelo da rede

Para a realização dos testes, utilizamos as redes neurais *ResNet 50*, *EfficientNet B2* e desenvolvemos um *Ensemble* combinando ambas. A *ResNet* selecionada (*ResNet 50*), possui 50 camadas do tipo residual e 23.538.690 parâmetros treináveis, enquanto que a *EfficientNet* selecionada (*EfficientNet B2*), possui 7.703.812 de parâmetros treináveis. Nos testes realizados com ambos os modelos, foram carregados os pesos pré-treinados com o dataset *ImageNet* e o otimizador escolhido foi o *Stochastic gradient descent* (SGD). Nas seções subsequentes detalharemos os modelos utilizados. Em ambos os casos, adicionamos à saída do modelo pré-treinado uma camada de *GlobalAveragePooling2D* e uma camada densa com ativação *sigmoid*, no intuito de realizar a distinção entre lente e não lente. Por fim, implementamos um *Ensemble* com os dois modelos, realizando uma média simples entre suas previsões.

4.1.1. ResNet

A rede neural residual profunda (*ResNet*) surgiu na edição de 2015 do ILSVRC, considerando as lições aprendidas nos experimentos de edições anteriores. As redes neurais com maior profundidade tinham grande potencial para aprender as características, porém vinham acompanhadas de problemas durante o treinamento. Uma classe destes problemas, *vanishing/exploding gradient*, poderia ser parcialmente tratado normalizando a inicialização da rede e utilizando camadas intermediárias de normalização [22]. Uma classe destes problemas, a saturação da acurácia, e os chamados *vanishing gradients* poderiam ser tratados com camadas residuais que compõe a *ResNet*[22]. A solução proposta consistia no uso de *shortcut connection*, onde uma camada pula uma ou mais camadas da rede e transmite/adiciona o seu re-

sultado nas camadas adiante. Essa técnica é utilizada para armazenar o mapeamento da identidade na camada de entrada e transmitir esses valores na saída das camadas seguintes, conforme figura 3. O desenvolvimento desse novo modelo, possibilitou a criação de uma rede com 152 camadas e desempenho superior ao alcançado pelas redes VGG [86]. Esse modelo ficou com a primeira colocação em diversas categorias na edição de 2015 do ILSVRC [22].

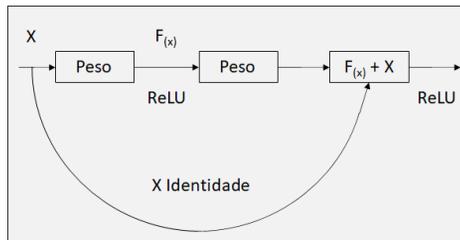


Figura 3: Esquema de Bloco identidade (bloco residual) contendo um atalho entre camadas (*Skip connection*).

4.1.2. EfficientNet

A rede neural *EfficientNet* foi proposta com objetivo de ganhar performance considerando uso eficiente de operações de ponto flutuante. Análises realizadas com essas redes demonstraram que era possível melhorar os resultados obtidos realizando o escalonamento de suas dimensões. Esse escalonamento consiste no aumento da profundidade [22] ou da largura [94], porém nesses casos apenas uma das dimensões era alterada [32]. O aumento de profundidade na *ResNet*, por exemplo, é eficiente no aumento da acurácia do modelo. Zagoruyko e Komodakis [94] realizaram experimentos que demonstram o grande benefício em se aumentar a largura da rede, ao invés da sua profundidade. Obtiveram excelentes resultados com uma rede "larga" de apenas 16 camadas. Este tipo de resultados motivou a explorar mais detalhada do escalonamento das CNNs. A proposta da *EfficientNet* é aperfeiçoar os resultados obtidos pelas CNNs, realizando o escalonamento balanceado de todas as 3 dimensões (profundidade, largura e resolução), conforme figura 4.

O método do balanceamento composto consiste em alterar todas as dimensões da rede a uma taxa constante [32]. Para a *EfficientNet* foi desenvolvido um novo modelo base, partindo de uma *MobileNet* e um método de busca de arquitetura neural (*Neural architecture search*, NAS) [95, 96], criando o modelo *EfficientNet B0*. Utilizando o método de escalonamento composto, foi criada uma família de modelos *EfficientNet* (*B1*, *B2*, *B3*, *B4*, *B5*, *B6* e *B7*). Estes modelos foram originados pelo conceito de escalonamento composto utilizando a mesma taxa de variação para cada dimensão, aumentando esses valores a cada nova versão [32].

4.1.3. Ensemble

O *ensemble* elaborado para esse trabalho, realiza o cálculo da média aritmética simples entre as previsões de cada rede,

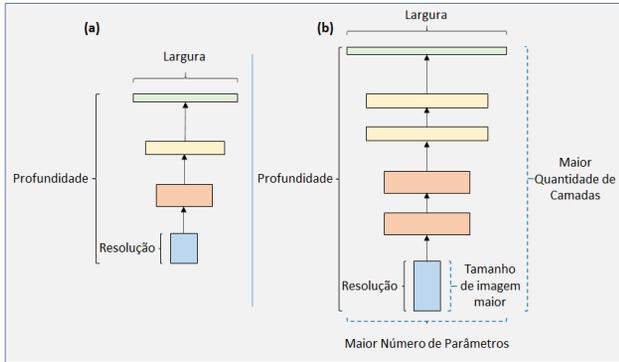


Figura 4: Modelo de Escalonamento Composto: (a) exemplo de uma rede genérica; (b) representa o método de escalonamento composto. Esse método escalona todas as dimensões uniformemente a uma taxa fixa.

de acordo a equação 5, sendo p é a probabilidade de que a imagem analisada contenha uma lente, em cada método:

$$P_{ensemble} = \frac{P_{resnet} + P_{efficientnet}}{2}. \quad (5)$$

Dado um mesmo conjunto de dados de teste, realizam-se as avaliações de cada uma das redes, retornando um banco de dados com a probabilidade de pertencimento à classe de interesse (lente ou não-lente), isto é, as predições. Em seguida, organizam-se as predições de cada rede para cada item e retira-se a média aritmética simples das probabilidades.

4.2. Métricas de avaliação dos resultados

4.2.1. Receiver operating characteristic curve

A curva ROC (*receiver operating characteristic curve*) é um método tradicional de avaliação para algoritmos de classificação em diferentes limiares (*thresholds*). Consiste na elaboração de um gráfico entre a taxa de positivos verdadeiros (TPR) e a taxa de falsos positivos (FPR), assumindo diferentes limiares de classificação. A taxa de positivos verdadeiros é dada pela razão

$$TPR = \frac{\text{n}^\circ \text{ de objetos corretamente classificados}}{\text{n}^\circ \text{ verdadeiro de objetos presentes}}, \quad (6)$$

enquanto a de falsos positivos é descrita pelo quociente [97]

$$FPR = \frac{\text{n}^\circ \text{ de casos negativos de objetos classificados}}{\text{n}^\circ \text{ verdadeiro de casos negativos de objetos}}. \quad (7)$$

O cálculo destes quocientes depende de uma definição de um limiar a partir do qual consideramos que o objeto foi classificado como lente ou não lente. Portanto, ao variar esse limiar entre 0 e 1, produzimos a referida curva. Uma maneira de analisar a qualidade da curva ROC é calculando a AUC (*area under curve*), isto é, a área sob a curva ROC. A predição aleatória em uma classificação binária produz uma AUC de 0.5, não importando quão desbalanceados estejam os dados, o que faz da AUC uma métrica de avaliação melhor do que a acurácia [97].

5. DATASET E DEFINIÇÃO DAS AMOSTRAS

A base de dados utilizada foi o conjunto de imagens da primeira edição do *Gravitational Lens Finding Challenge*. O *dataset* é constituído por 20.000 recortes 101×101 pixels de imagens de galáxias em diferentes desvios pro vermelho, nas bandas (u,g,r,i), simuladas de acordo observações *ground-based* (ver figura 5).

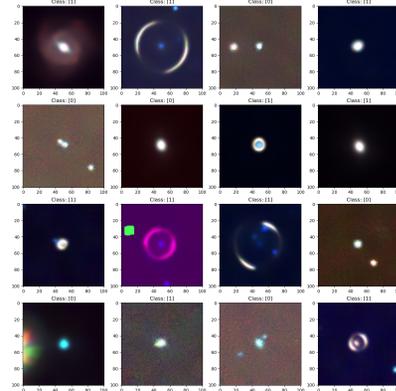


Figura 5: Grupo de algumas imagens presentes no *dataset* do *Gravitational Lens Finding Challenge*. As imagens tiveram os seus 3 canais sobrepostos e convertidos para o formato RGB, a fim de facilitar a visualização.

Quanto ao balanceamento das classes, o *dataset* se mostrou razoavelmente balanceado. Entre as 20.000 imagens, existem 8.021 em que o efeito forte de lente do tipo galáxia-galáxia está presente, e 11.979 imagens em que esse efeito não é verificado. As imagens foram geradas no formato FITS, de acordo com um algoritmo simulador de imagem, que reproduz a performance e os ruídos observados em diversos instrumentos ópticos [98], em conjunto a um simulador de imagens de galáxias, que reproduz o efeito forte de lente (para uma descrição extensa do dataset simulado veja [24]). A simulação ocorreu como se os objetos observados estivessem dentro de um cone de luz de 1.6 deg^2 que se estende até o desvio para o vermelho $z = 6$, incluindo o efeito de *halos* de matéria escura. Para simular as fontes lenteadas no plano de fundo das imagens, foram utilizados os objetos do *Hubble Ultra Deep Field* [24], que consiste em um catálogo de objetos retirados a partir de uma exposição de uma região de cerca de 11 arcmin^2 do céu, que conta com um conjunto de galáxias em desvios para o vermelho entre $z = 4$ e $z = 7$.

Por terem sido baseadas em *surveys ground-based*, as imagens apresentam bastante ruído, borramento e outros efeitos devido a interferência atmosférica. Borrões, flashes e outros artefatos, comuns em imagens CCDs, também se fazem presentes em alguns casos, que inviabiliza algumas imagens de serem analisadas pela inspeção visual, potencialmente confundindo também o algoritmo.

6. RESULTADOS

Na execução dos testes foi utilizada uma máquina com 125 GB de memória RAM, duas CPUs AMD Opteron(tm)

Processor 6370P e duas GPUs GeForce RTX 2080 Ti. Realizamos o treinamento com 50 épocas e *batch size* de 64, utilizando a validação cruzada com $K = 10$ *folde*s. Utilizamos os seguintes tipos de aumentação de dados: *rotation range=90*, *zca whitening*, *horizontal flip* e *vertical flip*. Executamos testes com as seguintes quantidades de dados de treino: 600, 550, 500, 450, 400, 350, 300, 250, 200, 100 e 50. Em seguida, realizamos uma segunda avaliação usando 100, 90, 80, 70, 60, 50, 40, 30 e 20 dados de treino.

O conjunto completo de dados foi dividido em três grupos, o primeiro consiste em 2.000 imagens, 10% do total de dados, que serão utilizados para validação do treino. Esse conjunto não será utilizado apenas como uma checagem de consistência no treinamento e avaliação de possível sobreajuste. O conjunto de treino foi utilizado efetivamente para otimização dos pesos da rede a partir da minimização da função custo (*Loss*). Cada treino foi realizado por 10 conjuntos distintos de treino, sorteados do total de 20.000 amostras. Por fim, o conjunto de teste representa todos os dados que não foram utilizados em nenhum dos 10 conjuntos de treino e nem na validação. Todos os resultados apresentados nos gráficos seguir utilizarão esse grupo. As medianas e os erros obtidos são provenientes do uso das 10 redes treinadas em conjuntos diferentes sendo aplicadas nos conjuntos de teste. A escolha desta estratégia ao invés da tradicional validação cruzada se dá em razão da abundância de dados se comparada ao tamanho dos conjuntos de treino. O tamanho do conjunto de teste pode ser dado pela seguinte relação:

$$Teste = C_{dados} - (10 \times Treino) - Val, \quad (8)$$

onde *Teste* representa o tamanho do conjunto de dados para teste, C_{dados} o conjunto completo de dados, *Val* o tamanho do conjunto de validação e *Treino* o conjunto de treinamento. Este último será redefinido 10 vezes. Por exemplo, com 600 dados de treino, teremos:

$$Teste = 20.000 - (10 \times 600) - 2.000. \quad (9)$$

Resultando em:

- Teste = 12.000
- Treino = 6.000
- Val = 2.000

6.1. Performance das redes

6.1.1. ResNet 50

Analisando os resultados do teste inicial com a ResNet, apresentados na figura 6 para o conjunto de teste, verificamos que a rede obteve uma AUC de 0,877 usando apenas 250 imagens de treino. A seguir, é possível notar que abaixo de 200 imagens de treinamento, a rede apresenta uma barra de erro significativamente maior. Isto aponta uma instabilidade do resultado, além do seu valor numérico ser inferior àquele obtido com 250 imagens. Na 7 apresentamos a ROC mediana

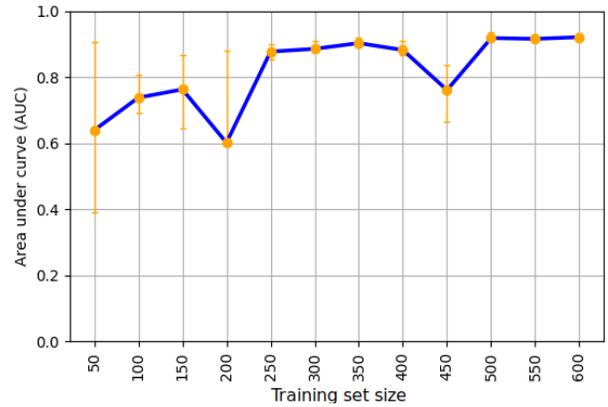


Figura 6: Resultados da AUC da curva ROC para diferentes tamanhos de grupos de treinamento, utilizando a ResNet 50.

no conjunto de teste com 250 imagens de treino os respectivos desvios. Observamos valores instáveis nos testes com 450 imagens, mesmo repetindo o teste diversas vezes. Analisamos a possibilidade do método ter alguma sensibilidade em termos de estabilidade no treino. Por isso, avaliamos o *batch size*, alterando o seu valor para 16, 32, 128 e 256 no teste com 450 imagens.

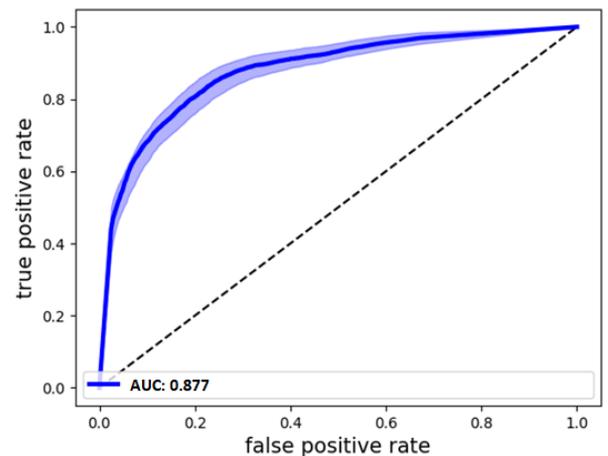


Figura 7: ROC mediana no conjunto de teste utilizando ResNet50 e 250 imagens de treino.

A figura 8 sugere que resultados com maior instabilidade no treino, para o conjunto fixo de 50 épocas, podem se beneficiar de um *batch* maior, ainda que as redes tenham convergido em menos de 50 épocas no treinamento.

Partindo dos resultados obtidos, realizamos um novo teste com *batch* imagens, obtendo os resultados apresentados na figura 9. Constatamos que as diferenças na performance do teste podem ter sido influenciadas pelo tamanho do *batch*. O melhor resultado alcançado nesta execução utilizou 200 imagens de treinamento e obteve AUC de 0,830. Porém, com esse tamanho de *batch*, alguns valores também apresentam variações numericamente negativas se comparadas com o resultado anterior de menor *batch*, mas consistentes dentro de 1σ especialmente com 150 imagens ou menos.

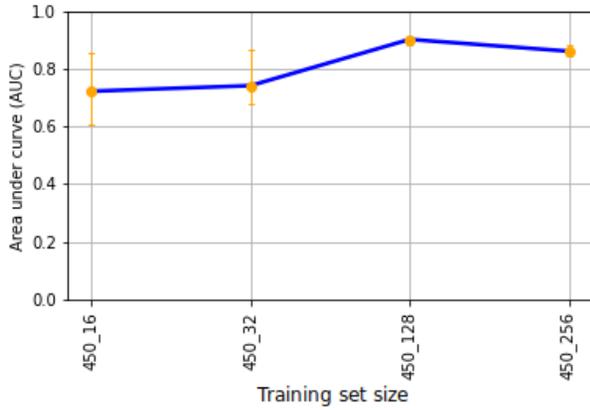


Figura 8: Resultados da AUC da curva ROC para diferentes tamanhos de Batch com 450 imagens de treino.

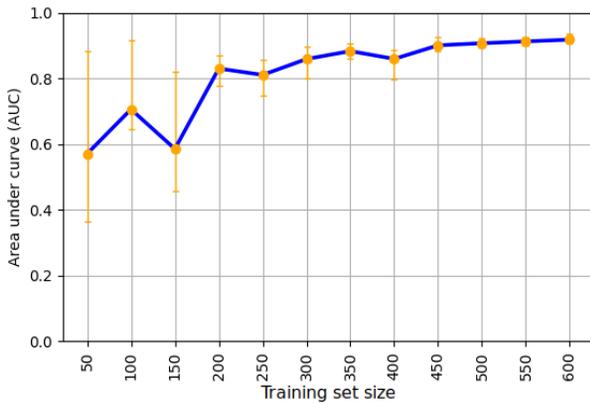


Figura 9: Resultados da AUC da curva ROC para diferentes tamanhos de grupos de treinamento, utilizando a ResNet 50 utilizando um *batch* de 128.

6.1.2. EfficientNet B2

Realizamos o teste inicial com a EfficientNet, obtendo os resultados apresentados na figura 10. Observamos que a rede obteve bom desempenho com 150 imagens de treino, apresentando AUC de 0,830. Os resultados apontam que este modelo ajustou bem no problema apresentado, visto a constância dos valores obtidos na AUC e as pequenas margens de erro, na maioria dos casos.

Na execução do teste com 150 imagens a AUC/ROC média das execuções das 10 *folds* é representada na figura 11.

6.2. Ensemble de Modelos

A partir do desempenho dos modelos apresentados anteriormente, focaremos nossas análises nos resultados dos testes obtidos nas faixas de 250, 150 e 100 imagens.

Realizamos testes com o Ensemble conforme descrito na seção 6.2. Os resultados estão apresentados na figura 12, apresentando AUC de 0,874 com 100 imagens de treino. Um valor superior àquele obtido pelas duas redes individualmente.

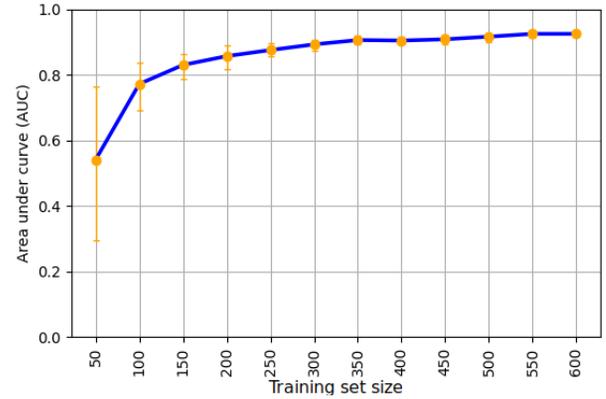


Figura 10: Resultados da AUC da curva ROC para diferentes tamanhos de grupos de treinamento, utilizando a EfficientNet B2.

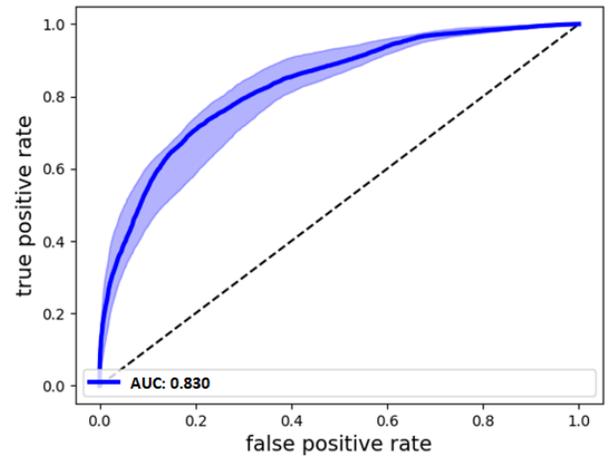


Figura 11: ROC mediana no conjunto de teste utilizando EfficientNet B2 e 250 imagens de treino.

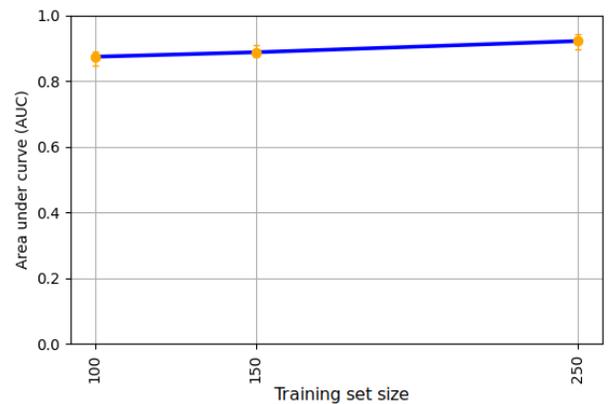


Figura 12: Resultados da AUC da curva ROC para diferentes tamanhos de grupos de treinamento, utilizando um Ensemble de redes.

A partir disso, analisamos conjuntos ainda menores, a fim de evidenciar o ponto no qual o Ensemble apresentará instabilidade no resultado. Apresentamos essa análise na figura 13, verificando que com 80 imagens de treinamento foi possível obter uma AUC de 0,796.

Apresentamos na figura 14 a ROC mediana do conjunto de

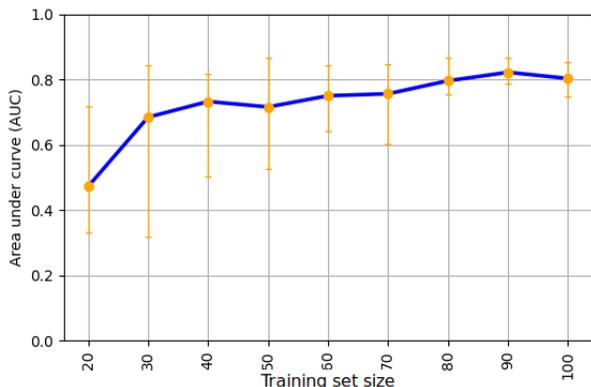


Figura 13: Resultados da AUC da curva ROC para diferentes tamanhos de grupos de treinamento, utilizando um Ensemble com uma quantidade menor de dados.

teste e seus respectivos desvios para 80 objetos, resultante das 10 diferentes definições de treinamento das execuções. É possível notar que o Ensemble individualmente apresenta maiores incertezas, se comparadas com as ROCs apresentadas nos métodos anteriores. No entanto, é importante considerar que o conjunto de dados no qual o Ensemble performou, é de quase a metade do conjunto das redes anteriores, o que sugere que o Ensemble é um método relevante para este tipo de análise de poucos dados.

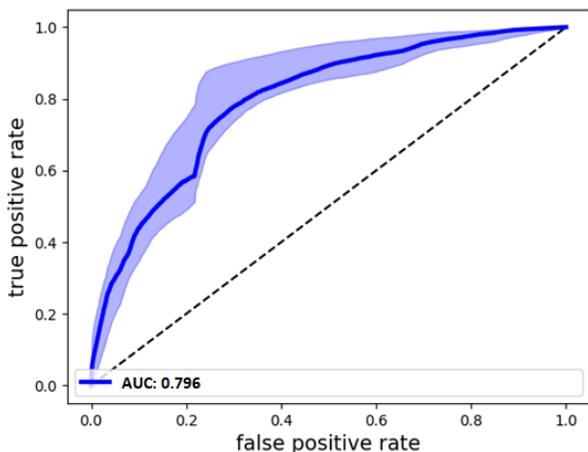


Figura 14: ROC mediana no conjunto de teste utilizando Ensemble de redes e 80 imagens de treino.

7. DISCUSSÃO

7.1. Sumário

Neste trabalho, analisamos duas diferentes arquiteturas, inspiradas nas soluções vencedoras das duas edições da competição de busca/classificação de Lentes Gravitacionais Fortes. Exploramos o regime de poucos dados, no âmbito da avaliação do quão estável é a generalização das redes, para uso futuro em datasets restritos e reais, que requeiram menor variedade para construção destes algoritmos. Pro-

pusemos um Ensemble de redes que, combinando o resultado das duas redes anteriores, conseguiu um desempenho na métrica AUC, superior ao obtido no treinamento das redes individualmente. Por conseguinte, obtivemos um desempenho de AUC de $\approx 0,8$ em um conjunto ainda menor de treino. Sumarizamos estes resultados nas tabelas I e II.

Tabela I: Resultados comparativos de AUCs de cada modelo.

Tamanho	<i>ResNet50</i>	<i>EfficientNet B2</i>	Ensemble
250	0,877	0,875	0,921
150	0,763	0,830	0,887
100	0,737	0,772	0,874

Tabela II: Resultados comparativos de AUCs do Ensemble.

Tamanho	AUC
100	0,803
90	0,823
80	0,796
70	0,756
60	0,750
50	0,716
40	0,732
30	0,684
20	0,472

7.2. O limite de poucos dados

O número de objetos necessários para treinar um algoritmo não depende somente da convergência da rede e sua estabilidade, como também de uma precisão aceitável. À medida em que necessitamos de um número menor de objetos, a expectativa é de que a rede generalize o conhecimento do que é uma lente/não lente, de maneira mais eficiente. Em aplicações com dados reais, a expectativa é de que sejam utilizados tantos sistemas, quanto possíveis. Em grandes surveys, estes números estão na ordem de centenas e, portanto, dentro do limite analisados nessa contribuição.

O modelo aqui proposto não requer que usemos cerca de 80 objetos para treino, isto é, cerca de 40 lentes reais para AUC de $\approx 0,8$. As escolhas e definições na construção de modelos que operem com tão poucos dados sugerem a construção de modelos cada vez mais eficientes. Estes, ainda poderiam, em princípio, ganhar performance num regime maior de dados.

7.3. Trabalhos futuros

Este trabalho apresenta um passo relevante na construção de redes eficientes na busca por Lentes gravitacionais. No entanto, existem várias possibilidades de investigações futuras. Por exemplo, o uso de outros modelos conhecidos de alta performance, como outras versões da Resnet e EfficientNet (B0 a B7, exceto a B2, testada no presente trabalho), além de outras propostas de rede como Inception [99] e Xception [100]. Outra abordagem interessante, seria investigar novas formas de construir o Ensemble, que

possam atribuir pesos dinâmicos a cada objeto testado na rede. Considerando os dados analisados, poderíamos testar a sua generalização em um domínio diferente, buscando outros datasets que tenham como origem algoritmos de simulação diferentes.

Agradecimentos

Os autores fizeram uso das máquinas de computação de alto desempenho Sci-Mind desenvolvidas e testadas para aplicações de Inteligência Artificial e gostariam de agradecer P. Russano e P. Souza Pereira pelo suporte e infraestrutura.

Referências Bibliográficas

- [1] José Ademir Sales Lima and Rose Clivia Santos. Do eclipse solar de 1919 ao espetáculo das lentes gravitacionais. *Revista Brasileira de Ensino de Física*, 41, 2019.
- [2] P. Schneider, J. Ehlers, and E.E. Falco. *Gravitational Lenses*. Astronomy and Astrophysics Library. Springer Berlin Heidelberg, 2013.
- [3] A.O. Petters, H. Levine, and J. Wambsganss. *Singularity Theory and Gravitational Lensing*. Progress in Mathematical Physics. Birkhäuser Boston, 2012.
- [4] S. Mollerach and E. Roulet. *Gravitational Lensing and Microlensing*. World Scientific, 2002.
- [5] Albert Einstein. Lens-like action of a star by the deviation of light in the gravitational field. *Science*, 84(2188):506–507, 1936.
- [6] Ronald Adler, Maurice Bazin, and Menahem Schiffer. *Introduction to general relativity*, volume 220. McGraw-Hill New York, 1975.
- [7] Lewis Ryder. *Introduction to general relativity*. Cambridge University Press, 2009.
- [8] Dennis Walsh, Robert F Carswell, and Ray J Weymann. 0957+ 561 a, b: twin quasistellar objects or gravitational lens? *Nature*, 279(5712):381–384, 1979.
- [9] RJ Weymann, FH Chaffee Jr, M Davis, NP Carleton, D Walsh, and RF Carswell. Multiple-mirror telescope observations of the twin qos 0957+ 561 a, b. *The Astrophysical Journal*, 233:L43–L46, 1979.
- [10] Scott Dodelson, Katrin Heitmann, Chris Hirata, Klaus Honscheid, Aaron Roodman, Uroš Seljak, Anže Slosar, and Mark Trodden. Cosmic visions dark energy: *Science*. *arXiv preprint arXiv:1604.07626*, 2016.
- [11] Željko Ivezić, Steven M Kahn, J Anthony Tyson, Bob Abel, Emily Acosta, Robyn Allsman, David Alonso, Yusra Al-Sayyad, Scott F Anderson, John Andrew, et al. Lsst: from science drivers to reference design and anticipated data products. *The Astrophysical Journal*, 873(2):111, 2019.
- [12] Paul A Abell, Julius Allison, Scott F Anderson, John R Andrew, J Roger P Angel, Lee Armus, David Arnett, SJ Asztalos, Tim S Axelrod, Stephen Bailey, et al. Lsst science book, version 2.0. *arXiv preprint arXiv:0912.0201*, 2009.
- [13] Thomas E Collett. The population of galaxy–galaxy strong lenses in forthcoming optical imaging surveys. *The Astrophysical Journal*, 811(1):20, 2015.
- [14] Anton T Jaelani, Ni Putu AP Emas, Premana W Premadi, and Fargiza AM Mulki. Constraints on dark energy models in cosmology from double-source plane strong lensing system. In *Journal of Physics: Conference Series*, volume 1245, page 012017. IOP Publishing, 2019.
- [15] Masayuki Tanaka, Kenneth C Wong, Anupreeta More, Arsha Dezuka, Eiichi Egami, Masamune Oguri, Sherry H Suyu, Alessandro Sonnenfeld, Ryo Higuchi, Yutaka Komiyama, et al. A spectroscopically confirmed double source plane lens system in the hyper supprime-cam subaru strategic program. *The Astrophysical Journal Letters*, 826(2):L19, 2016.
- [16] Ran Li, Carlos S Frenk, Shaun Cole, Liang Gao, Sownak Bose, and Wojciech A Hellwing. Constraints on the identity of the dark matter from strong gravitational lenses. *Monthly Notices of the Royal Astronomical Society*, 460(1):363–372, 2016.
- [17] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.
- [18] R Christopher Lacher, Pamela K Coats, Shanker C Sharma, and L Franklin Fant. A neural network for classifying the financial health of a firm. *European Journal of Operational Research*, 85(1):53–65, 1995.
- [19] Jouko Lampinen, Seppo Smolander, and Markku Korhonen. Wood surface inspection system based on generic visual features. In *Industrial Applications of Neural Networks*, pages 35–42. World Scientific, 1998.
- [20] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*, 2014.
- [21] Le Cun Yan, B Yoshua, and H Geoffrey. Deep learning. *nature*, 521(7553):436–444, 2015.
- [22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [23] Clecio Bom, Jason Poh, Brian Nord, Manuel Blanco-Valentin, and Luciana Dias. Deep learning in wide-field surveys: Fast analysis of strong lenses in ground-based cosmic experiments. *arXiv preprint arXiv:1911.06341*, 2019.
- [24] R Benton Metcalf, M Meneghetti, Camille Avestruz, Fabio Bellagamba, Clécio R Bom, Emmanuel Bertin, Rémi Cabanac, F Courbin, Andrew Davies, Etienne Decencière, et al. The strong gravitational lens finding challenge. *Astronomy & Astrophysics*, 625:A119, 2019.
- [25] CR Bom, M Makler, MP Albuquerque, and CH Brandt. A neural network gravitational arc finder based on the mediatrix filamentation method. *Astronomy & Astrophysics*, 597:A135, 2017.
- [26] François Lanusse, Quanbin Ma, Nan Li, Thomas E Collett, Chun-Liang Li, Siamak Ravanbakhsh, Rachel Mandelbaum, and Barnabás Póczos. Cmu deeplens: deep learning for automatic image-based galaxy–galaxy strong lens finding. *Monthly Notices of the Royal Astronomical Society*, 473(3):3895–3906, 2018.
- [27] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.
- [28] Zhiang Hu. Real-time head pose estimation with convolutional neural networks. Lee, K.-F., *Automatic Speech Recognition: The Development of the SPHINX SYSTEM*, Kluwer Academic Publishers, Boston, 1989.
- [29] Dan C Cireşan, Ueli Meier, Jonathan Masci, Luca M Gambardella, and Jürgen Schmidhuber. High-performance neural networks for visual object classification. *arXiv preprint arXiv:1102.0183*, 2011.
- [30] Yann Le Cun and Yoshua Bengio. Word-level training of a handwritten word recognizer based on convolutional neural networks. In *Proceedings of the 12th IAPR International*

- Conference on Pattern Recognition, vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5), volume 2, pages 88–92. IEEE, 1994.
- [31] AD Dongare, RR Kharde, and Amit D Kachare. Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology (IJEIT)*, 2(1):189–194, 2012.
- [32] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, 2019.
- [33] Matthias Bartelmann and Peter Schneider. Weak gravitational lensing. *Physics Reports*, 340(4-5):291–472, 2001.
- [34] Ian Morison. *Introduction to astronomy and cosmology*. John Wiley & Sons, 2013.
- [35] Barbara Ryden. *Introduction to cosmology*. Cambridge University Press, 2017.
- [36] Matts Roos. *Introduction to cosmology*. John Wiley & Sons, 2015.
- [37] Martin Makler. Lentas gravitacionais: Um rico recurso didático na ocasião da celebração dos 100 anos do eclipse de sobral. *Conexões-Ciência e Tecnologia*, 13(2):21–36, 2019.
- [38] Ramesh Narayan and Matthias Bartelmann. Lectures on gravitational lensing. *arXiv preprint astro-ph/9606001*, 1996.
- [39] CS Kochanek, CR Keeton, and BA McLeod. The importance of einstein rings. *The Astrophysical Journal*, 547(1):50, 2001.
- [40] Clecio Roque De Bom, Martin Makler, and Marcelo Portes de Albuquerque. Mediatrix method for filamentation of objects in images. *NOTAS TÉCNICAS*, 6(2), 2017.
- [41] Clecio Roque De Bom, Martin Makler, and Marcelo Albuquerque. Método de reconhecimento de arcos gravitacionais com redes neurais utilizando a decomposição por mediatrizes. *NOTAS TÉCNICAS*, 2(1), 2012.
- [42] Cristina Furlanetto, Basilio X Santiago, Martin Makler, Clécio de Bom, Carlos H Brandt, Angelo Fausti Neto, Pedro C Ferreira, Luiz Nicolaci da Costa, and Marcio AG Maia. A simple prescription for simulating and characterizing gravitational arcs. *Astronomy & Astrophysics*, 549:A80, 2013.
- [43] J Estrada, J Annis, HT Diehl, PB Hall, T Las, H Lin, M Makler, KW Merritt, V Scarpine, S Allam, et al. A systematic search for high surface brightness giant arcs in a sloan digital sky survey cluster sample. *The Astrophysical Journal*, 660(2):1176, 2007.
- [44] Clecio Roque De Bom, Marcelo Portes de Albuquerque, and Martin Makler. Uso da transformada de hough para detecção de arcos. *NOTAS TÉCNICAS*, 2(1), 2012.
- [45] EE Falco, CS Kochanek, J Lehár, BA McLeod, JA Munoz, CD Impey, C Keeton, CY Peng, and H-W Rix. The castles gravitational lensing tool. In *Gravitational Lensing: Recent Progress and Future Go*, volume 237, page 25, 2001.
- [46] Sjur Refsdal and Jean Surdej. Gravitational lenses. *Reports on Progress in Physics*, 57(2):117, 1994.
- [47] Fritz Zwicky. Nebulae as gravitational lenses. *Physical Review*, 51(4):290, 1937.
- [48] Geneviève Soucail. Gravitational lensing: Recent progress and future goals. In *AIP Conference Proceedings*, volume 586, pages 233–244. American Institute of Physics, 2001.
- [49] H. Ebeling, M. Stockmann, J. Richard, J. Zabl, G. Brammer, S. Toft, and A. Man. Thirty-fold: Extreme Gravitational Lensing of a Quiescent Galaxy at $z = 1.6$. *Astrophys. J.*, 852:L7, Jan 2018.
- [50] J. Richard, T. Jones, R. Ellis, D. P. Stark, R. Livermore, and M. Swinbank. The emission line properties of gravitationally lensed $< 1.5 < z < 5$ galaxies. *Monthly Notices of the Royal Astronomical Society*, 413:643–658, May 2011.
- [51] T. A. Jones, A. M. Swinbank, R. S. Ellis, J. Richard, and D. P. Stark. Resolved spectroscopy of gravitationally lensed galaxies: recovering coherent velocity fields in subluminal $z \sim 2-3$ galaxies. *Monthly Notices of the Royal Astronomical Society*, 404:1247–1262, May 2010.
- [52] Takahiro Sumi, K Kamiya, DP Bennett, IA Bond, F Abe, CS Botzler, A Fukui, K Furusawa, JB Hearnshaw, Y Itow, et al. Unbound or distant planetary mass population detected by gravitational microlensing. *Nature*, 473(7347):349, 2011.
- [53] Alexander Zakharov. Lensing by exotic objects. *General Relativity and Gravitation*, 42(9):2301–2322, 2010.
- [54] Donald G York, J Adelman, John E Anderson Jr, Scott F Anderson, James Annis, Neta A Bahcall, JA Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, et al. The sloan digital sky survey: Technical summary. *The Astronomical Journal*, 120(3):1579, 2000.
- [55] V Belokurov, NW Evans, PC Hewett, A Moiseev, RG McMahon, SF Sanchez, and LJ King. Two new large-separation gravitational lenses from sdss. *Monthly Notices of the Royal Astronomical Society*, 392(1):104–112, 2009.
- [56] Jeffrey M Kubo, Sahar S Allam, Emily Drabek, Huan Lin, Douglas Tucker, Elizabeth J Buckley-Geer, H Thomas Diehl, Marcelle Soares-Santos, Jiangang Hao, Matthew Wiesner, et al. The sloan bright arcs survey: discovery of seven new strongly lensed galaxies from $z = 0.66-2.94$. *The Astrophysical Journal Letters*, 724(2):L137, 2010.
- [57] JP Willis, PC Hewett, SJ Warren, S Dye, and N Maddox. The ols-lens survey: the discovery of five new galaxy–galaxy strong lenses from the sdss. *Monthly Notices of the Royal Astronomical Society*, 369(3):1521–1528, 2006.
- [58] IWA Browne, PN Wilkinson, NJF Jackson, ST Myers, CD Fassnacht, LVE Koopmans, DR Marlow, M Norbury, D Rusin, CM Sykes, et al. The cosmic lens all-sky survey-ii. gravitational lens candidate selection and follow-up. *Monthly Notices of the Royal Astronomical Society*, 341(1):13–32, 2003.
- [59] Jelte TA de Jong, Gijs A Verdoes Kleijn, Konrad H Kuijken, and Edwin A Valentijn. The kilo-degree survey. *Experimental Astronomy*, 35(1):25–44, 2013.
- [60] CE Petrillo, CRESCENZO Tortora, S Chatterjee, G Vernardos, LVE Koopmans, G Verdoes Kleijn, NICOLA ROSARIO Napolitano, G Covone, P Schneider, ANIELLO Grado, et al. Finding strong gravitational lenses in the kilo degree survey with convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 472(1):1129–1150, 2017.
- [61] Dark Energy Survey Collaboration et al. The dark energy survey. *arXiv preprint astro-ph/0510346*, 2005.
- [62] B Nord, E Buckley-Geer, H Lin, HT Diehl, J Helsby, N Kuropatkin, Adam Amara, T Collett, S Allam, GB Caminha, et al. Observation and confirmation of six strong-lensing systems in the dark energy survey science verification data. *The Astrophysical Journal*, 827(1):51, 2016.
- [63] Colin Jacobs, Thomas Collett, K Glazebrook, C McCarthy, AK Qin, TMC Abbott, FB Abdalla, J Annis, S Avila, K Bechtol, et al. Finding high-redshift strong lenses in des using convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 484(4):5330–5349, 2019.
- [64] H Thomas Diehl, EJ Buckley-Geer, KA Lindgren, B Nord, H Gaitsch, S Gaitsch, H Lin, S Allam, TE Collett, C Furlanetto, et al. The des bright arcs survey: hundreds of candidate strongly lensed galaxy systems from the dark energy survey science verification and year 1 observations. *The Astrophysical Journal Supplement Series*, 232(1):15, 2017.
- [65] Mirjana Pović, Isabel Márquez, Josefa Masegosa, Jaime Perea, A del Olmo, C Simpson, José Alfonso López

- Aguerri, Begoña Ascaso, Yolanda Jiménez-Teja, Carlos López-Sanjuán, et al. The impact from survey depth and resolution on the morphological classification of galaxies. *Monthly Notices of the Royal Astronomical Society*, 453(2):1644–1668, 2015.
- [66] Leonidas A Moustakas, Phil Marshall, Jeffrey A Newman, Alison L Coil, Michael C Cooper, Marc Davis, Christopher D Fassnacht, Puragra Guhathakurta, Andrew Hopkins, Anton Koekemoer, et al. A strong-lens survey in aegis: The influence of large-scale structure. *The Astrophysical Journal Letters*, 660(1):L31, 2007.
- [67] CE Petrillo, CRESCENZO Tortora, Giorgos Varnardos, LVE Koopmans, G Verdoes Kleijn, M Bilicki, NICOLA ROSARIO Napolitano, Saikat Chatterjee, G Covone, Andrej Dvornik, et al. Links: discovering galaxy-scale strong lenses in the kilo-degree survey using convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 484(3):3879–3896, 2019.
- [68] C Schaefer, M Geiger, T Kuntzer, and J-P Kneib. Deep convolutional neural networks as strong gravitational lens detectors. *Astronomy & Astrophysics*, 611:A2, 2018.
- [69] Xiaosheng Huang, Christopher Storfer, V Ravi, A Pilon, M Domingo, DJ Schlegel, S Bailey, A Dey, RR Gupta, D Herrera, et al. Finding strong gravitational lenses in the desi decam legacy survey. *The Astrophysical Journal*, 894(1):78, 2020.
- [70] Xiaosheng Huang, Christopher Storfer, A Gu, V Ravi, A Pilon, W Sheu, R Venguswamy, S Banka, A Dey, M Landriau, et al. Discovering new strong gravitational lenses in the desi legacy imaging surveys. *The Astrophysical Journal*, 909(1):27, 2021.
- [71] Jeff Heaton. *Artificial Intelligence for Humans. Volume 3: Deep Learning and Neural networks*. CreateSpace Independent Publishing: North Charleston, SC, USA, 2015.
- [72] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.
- [73] Pavel Hála. Spectral classification using convolutional neural networks. *arXiv preprint arXiv:1412.8341*, 2014.
- [74] Sander Dieleman, Kyle W Willett, and Joni Dambre. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Monthly notices of the royal astronomical society*, 450(2):1441–1459, 2015.
- [75] Diego Tuccillo, Etienne Decencière, Santiago Velasco-Forero, et al. Deep learning for studies of galaxy morphology. *Proceedings of the International Astronomical Union*, 12(S325):191–196, 2016.
- [76] Edward J Kim and Robert J Brunner. Star-galaxy classification using deep convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, page stw2672, 2016.
- [77] Ben Hoyle. Measuring photometric redshifts using galaxy images and deep neural networks. *Astronomy and Computing*, 16:34–40, 2016.
- [78] Zizhao He, Xinzhong Er, Qian Long, Dezi Liu, Xiangkun Liu, Ziwei Li, Yun Liu, Wenqiang Deng, and Zuhui Fan. Deep learning for strong lensing search: tests of the convolutional neural networks and new candidates from kids dr3. *Monthly Notices of the Royal Astronomical Society*, 497(1):556–571, 2020.
- [79] R Li, NR Napolitano, C Tortora, C Spiniello, LVE Koopmans, Z Huang, N Roy, G Varnardos, S Chatterjee, B Giblin, et al. New high-quality strong lens candidates with deep learning in the kilo-degree survey. *The Astrophysical Journal*, 899(1):30, 2020.
- [80] Chris Lintott, Kevin Schawinski, Steven Bamford, Anže Slosar, Kate Land, Daniel Thomas, Edd Edmondson, Karen Masters, Robert C. Nichol, M. Jordan Raddick, Alex Szalay, Dan Andreescu, Phil Murray, and Jan Vandenberg. Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies*. *Monthly Notices of the Royal Astronomical Society*, 410(1):166–178, 12 2010.
- [81] Mike Walmsley, Chris Lintott, Tobias Geron, Sandor Kruk, Coleman Krawczyk, Kyle W Willett, Steven Bamford, William Keel, Lee S Kelvin, Lucy Fortson, et al. Galaxy zoo decals: Detailed visual morphology measurements from volunteers and deep learning for 314,000 galaxies. *arXiv preprint arXiv:2102.08414*, 2021.
- [82] J Amiaux, R Scaramella, Y Mellier, B Altieri, C Burigana, Antonio Da Silva, P Gomez, J Hoar, R Laureijs, E Maiorano, et al. Euclid mission: building of a reference survey. In *Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave*, volume 8442, page 84420Z. International Society for Optics and Photonics, 2012.
- [83] Jeffrey Dean, Greg S Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Quoc V Le, Mark Z Mao, Marc’Aurelio Ranzato, Andrew Senior, Paul Tucker, et al. Large scale distributed deep networks. 2012.
- [84] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [85] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.
- [86] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [87] George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- [88] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Metasgd: Learning to learn quickly for few-shot learning. *arXiv preprint arXiv:1707.09835*, 2017.
- [89] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Computing Surveys (CSUR)*, 53(3):1–34, 2020.
- [90] Qianru Sun, Yaoyao Liu, Tat-Seng Chua, and Bernt Schiele. Meta-transfer learning for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 403–412, 2019.
- [91] Wang Yan, Jordan Yap, and Greg Mori. Multi-task transfer methods to improve one-shot learning for multimedia event detection. In *BMVC*, pages 37–1, 2015.
- [92] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. *arXiv preprint arXiv:1606.04080*, 2016.
- [93] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. 2016.
- [94] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.
- [95] Han Cai, Ligeng Zhu, and Song Han. Proxylessnas: Direct neural architecture search on target task and hardware, 2019.
- [96] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, Mark Sandler, Andrew Howard, and Quoc V. Le. Mnasnet: Platform-aware neural architecture search for mobile, 2019.
- [97] Andreas C Müller and Sarah Guido. *Introduction to machine learning with Python: a guide for data scientists*. ”O’Reilly Media, Inc.”, 2016.
- [98] A Grazian, A Fontana, C De Santis, S Gallozzi, E Giallongo,

- and F Di Pangrazio. The large binocular camera image simulator. Publications of the Astronomical Society of the Pacific, 116(822):750, 2004.
- [99] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1–9, 2015.
- [100] François Chollet. Xception: Deep learning with depth-wise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1251–1258, 2017.