Estudo Comparativo entre Arquiteturas de Redes Neurais Profundas para Classificação de Imagens de Galáxias do Galaxy Zoo 2

Allan Maffra, Vitória Gabriely Sousa e Caio Rafael do Nascimento Santiago

September 15, 2023

Abstract

In the current literature there are Artificial Neural Network architectures developed for image classification, which it can serves as support and solves different problems in several areas; and over the years, it ended up becoming allies in carrying out processes, which it was being done manually. This project will show some current architectures of Neural Network of image classification, such as: Basic Convolutional Networks, Residual Networks, Inceptions Networks, Transformers Networks, their technologies and variations; using as a basis for this study, images of galaxies taken from the Galaxy Zoo 2 project, in order to compare the results obtained from each of the networks tested. Two types of samples were used, the original images, which it remained in its original state since the collection, and the same after a process of segmentation of the target galaxy, where it was submitted to an image treatment process to remove noise and objects. The separation of the images was performed based on the Hubble classification morphology, where it was grouped by labels and their similarities, being between Ellipticals (E) and Spirals (S).

Resumo

Na literatura atual existe arquiteturas de redes artificiais desenvolvidas para a classificação de imagens, nos quais conseguem servir como apoio e solucionar problemas diversos em várias áreas; e com o passar dos anos, acabaram se tornando aliadas na realização de processos, nos quais estavam sendo feitos de forma manual. Nesse projeto será mostrado algumas arquiteturas atuais de Redes Neurais de classificação de imagens, como: Redes Convolucionais básicas, Redes Residuais, Redes Inceptions, Redes Transformers, suas tecnologias e variações; sendo usando como base para esse estudo, imagens de galáxias retiradas do projeto Galaxy Zoo 2, com o objetivo de comparar os resultados obtidos de cada uma das redes testadas. Foram utilizados dois tipos de amostras, as imagens originais, nos quais permaneceram em seu estado original desde a coleta, e as mesmas após um processo de segmentação da galáxia alvo, onde foram submetidas a um processo de tratamento de imagens para a retirada de ruídos e objetos. A separação das imagens foi realizada com base na morfologia de classificação Hubble, onde foram agrupadas por rótulos e suas semelhanças, ficando entre Elípticas (E) e Espirais (S).

1 Introdução

Com o início da astronomia moderna, inúmeros estudos sobre objetos astronômicos foram possibilitados de avançarem, sendo, o mapeamento do espaço [Dunlop et al., 2004, Bunker et al., 2010, Jiang et al., 2021]. Com a ajuda de instrumentos de observação e a tecnologia em constante evolução, novos objetos astronômicos, como estrelas e galáxias, vêm sendo catalogados, e consequentemente, gerando uma grande quantidade de informações e amostras de imagens.

Um dos grandes desafios dessa área vem sendo a classificação desses objetos, já que devido ao grande número de dados coletados, o trabalho manual se tornou algo inviável para os grupos de pesquisadores [Xiao-Qing and Jin-Meng, 2021], criando-se projetos como o Galaxy Zoo [Bamford et al., 2009]. Nele, a ajuda de colaboradores, entre pesquisadores da área e o público em geral, está sendo usado como uma forma paliativa para atender a alta demanda na parte da classificação de imagens de galáxias, porém, por continuar sendo um processo manual, mesmo que realizado por um número maior de pessoas, problemas de precisão de classificação e de agilidade permaneceram, gerando-se a necessidade de um apoio tecnológico capaz de automatizar ou auxiliar o trabalho manual, padronizando e acelerando o processo Hart et al. [2016].

Com o desenvolvimento contínuo de algoritmos de Machine Learning e Deep Learning, nos quais conseguem processar grandes quantidades de dados, essa etapa está podendo receber um grande reforço, se tornando um importante aliado para a realização das classificações dessas imagens, ou até mesmo, serem automatizadas. Neste projeto os mais atuais métodos de classificações de imagens, suas características e os resultados obtidos em cima de uma amostra retirada através do Hubble Space Telescope, serão mostrados.

Neste projeto será apresentado um comparativo de resultados sob as classificações das galáxias através de redes neurais profundas atuais de diferentes arquiteturas, focadas nesse problema específico, como por exemplo: redes Inception Szegedy et al. [2015], redes ResNet He et al. [2016] e redes Transformer Dai et al. [2021]. Além disso, para gerar diferentes amostras de resultados serão utilizados métodos de tratamento de imagens e também, o Mecanismo de Atenção Bahdanau et al. [2014] como complemento na rede ResNet.

2 Dados

Os dados utilizados neste trabalho foram retirados diretamente do banco de dados do projeto Galaxy Zoo [Bamford et al., 2009], sendo ele, um projeto de ciência cidadã que fornece classificações morfológicas entre galáxias elípticas e espirais para quase um milhão de galáxias com a ajuda do público em geral [Barchi et al., 2020]. Porém, ao longo desse trabalho, foi utilizado as classificações morfológicas da segunda fase deste projeto, o Galaxy Zoo 2 [Hart et al., 2016], que fornece características mais detalhadas das galáxias, como quantidade de braços, barras, grau de arredondamento e muitos outros. Esses dados foram mesclados com dados do projeto intitulado como *Galaxy Zoo 2: Images from Original Sample* [Hart et al., 2016], onde além de fornecer as classificações morfológicas de forma tabulada dos mesmos dados do Galaxy Zoo 2, foi possível ter acesso as imagens das galáxias rotuladas. Com isso, foi possível ter uma

gama de dados para utilizar como treino, teste e validação das redes neurais estudadas nesse projeto.

3 Classificações Morfológicas

Antes de iniciar o estudo sobre as redes neurais, é necessário estudar e entender as classificações das galáxias. Embora possua muitos tipos de classificações sobre as galáxias ao decorrer da história, o esquema de classificação morfológica de galáxias mais conhecida é apresentado por Edwin Hubble Hubble [1926]. Primeiramente, seu estudo classifica as galáxias em quatro classes gerais: Elípticas (E), Espirais ordinais (S), Espirais barradas (SB) e Irregulares (I), onde, em cada uma delas ocorre subdivisões ao apresentar mais características, como: grau de arredondamento, número de braços, grau de enrolamento dos braços, barra central, centro proeminente etc. Sendo essas características, possíveis de apresentarem mais de quinhentas fusões diferentes no projeto Galaxy Zoo 2 Hart et al. [2016], nos quais podem definir um diferente tipo de galáxia.

4 Redes Neurais Artificiais

As Redes Neurais Artificiais foram amadurecendo através de novas pesquisas e estudos realizados ao passar dos anos, oferecendo um poderoso conjunto de ferramentas capaz de resolver problemas de reconhecimento de padrões, além de ser uma ótima ferramenta para a realização de processamentos de dados com uma alta velocidade de performance e aprendizado [Bishop, 1994]. Mas assim que começaram a ser mais exploradas, ficou evidente que a versão mais básica de redes neurais artificiais, a rede neural artificial profunda, não atendia todas as necessidades dos estudiosos da área, então foram desenvolvidos novos tipos de neurônios e estruturas mais complexas através dela, como as Redes Neurais Convolucionais (CNNs) LeCun et al. [1999], Redes Neurais Recorrentes (RNNs) [Jordan, 1986, Rumelhart et al., 1985].

4.1 Redes Neurais Convolucionais (CNNs) Básicas

As Redes Neurais Convolucionais surgiram como uma solução para o problema da alta densidade em redes com muitos parâmetros, mas causaram uma grande revolução na solução de problemas complexos como em imagens, áudio e texto [LeCun et al., 1999]. Convoluções são operadores lineares que agem como filtros sobre o sinal de entrada, e no caso de imagens já eram amplamente utilizadas para resolver problemas simples de detecção de bordas ou realce de contraste [Shapiro et al., 2001, Getreuer, 2013]. Porém, em redes neurais, essas camadas apresentam um significado ainda mais sofisticado, pois a composição de diversos filtros realizados sequencialmente ganham o sentido de detectar sinais mais complexos. Por fim, a camada de classificação tem o trabalho de atuar em um conjunto simplificado de características complexas, aumentando o poder de classificação das redes. Essa aplicação é comumente usada para a classificação de imagens e outros problemas com entradas complexas.

A AlexNet [Krizhevsky et al., 2012] é umas das mais populares redes neurais convolucionais, ela é estruturada em camadas, sendo elas, cinco convolucionais, intercaladas

por funções de *pooling*, e três camadas finais densamente conectadas. A primeira camada recebe entradas de tamanho 227x227x3, filtrando-as com 96 kernels de tamanho 11x11, normalizando e passando pela função de *pooling*. A segunda camada convolucional recebe como entrada a saída da primeira camada, passando por novos 256 filtros de tamanho 5x5, e novamente, sendo normalizada e passada pela função de *pooling*. As próximas três camadas seguem o mesmo processo, apenas mudando a quantidade de filtros e os tamanhos dos kernels utilizados, sendo eles: 384 filtros de tamanho 3x3, 384 filtros de tamanho 3x3 e 256 filtros de tamanho 3x3, sequencialmente. As camadas totalmente conectadas contêm 4096 neurônios cada, sendo utilizada a função de ativação *Rectified Linear Units* (ReLUs), para evitar problemas de *Vanishing Gradient* (VG), e por apresentar o melhor resultado referente a velocidade de aprendizagem comparado com as funções de sigmoide e tangente [Krizhevsky et al., 2012].

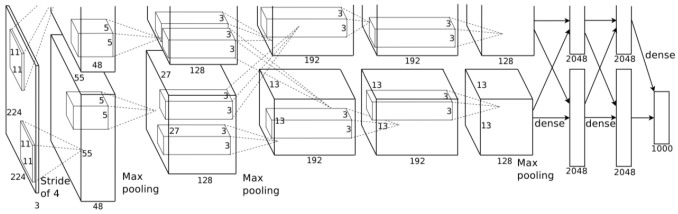

Figura 1: Alexnet

Por ter sido projetada para ser utilizada em aplicativos móveis, a MobileNet [Howard et al., 2017] tem um número reduzido de parâmetros devido ao processo de Convoluções Separadas em Profundidade, no qual ajuda a diminuir sua complexidade computacional, deixando a rede mais rápida e leve, comparado com redes nos quais usam convoluções tradicionais.

Esse conceito é composto por dois processos de convoluções: convolução em profundidade e convolução pontual. A convolução em profundidade trabalha usando filtros que diminuem a dimensão da imagem sem modificar a profundidade dos canais de entrada, já a convolução pontual, na qual trabalha com foco na profundidade de canais, utiliza um kernel fixado em 1x1 com a profundidade dependente do tamanho da profundidade da imagem de entrada, para gerar uma imagem de profundidade de tamanho 1 sem alteração das dimensões. Assim, pode ser gerado uma imagem com profundidade do tamanho desejado, como uma convolução normal.

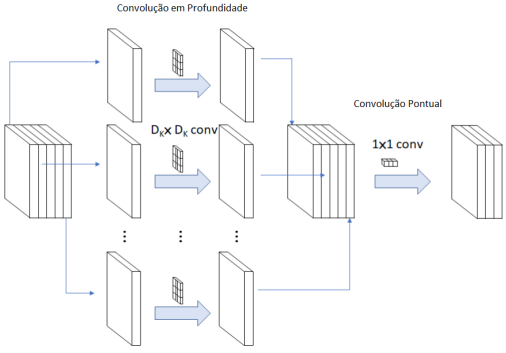

Figura 2: Convolução Separada em Profundidade

5 Inception

O módulo Inception [Szegedy et al., 2015] foi desenvolvido pela Google em parceria com várias universidades, entregando uma nova forma de organizar os filtros convolucionais por camadas, com o objetivo de diminuir o tamanho das redes e parâmetros que estavam ficando cada vez maiores e propensos a problemas como *Overfitting*. Diferente de redes anteriores, como a AlexNet [Krizhevsky et al., 2012] ou a VGGNet [Simonyan and Zisserman, 2014], onde a organização consistia em um tamanho único de filtro convolucional por camada, o módulo Inception recebe as informações das imagens no início de cada camada, processando-as por múltiplos filtros de diferentes tamanhos, no final da camada os resultados de cada filtro são concatenados. Em cada módulo da rede são agrupados filtros extratores de informações dos tamanhos 1x1, 3x3 e 5x5, mais um filtro *max pooling* 3x3 que resume as informações iniciais da camada. Com o objetivo de reduzir ainda mais as dimensões, foram inseridos filtros convolucionais 1x1 antes dos filtros 3x3, 5x5 e após o *max pooling* 3x3. A variação do tamanho do kernel dos filtros no módulo Inception tem como propósito obter diferentes tipos de características, quanto maior o kernel, mais características globais serão extraídas, por sua vez, quanto menor o kernel, mais características locais ou específicas serão extraídas.

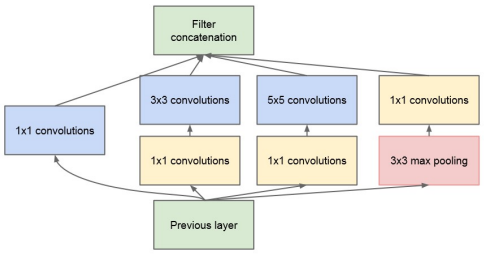

Figura 3: Módulo Inception

A GoogLeNet Szegedy et al. [2015] é uma rede que também utiliza de módulos Inception, vencedora da competição da classificação de imagens ImageNet Large Scale Visual Recognition (ILSVRC) em 2014 [Russakovsky et al., 2015] é composta por 22 camadas com 9 módulos Inception. A rede conta com a diminuição de parâmetros comparado com as redes anteriores, por conta disso, houve um aumento em sua largura devido ao acréscimo de filtros por módulo, como descritos anteriormente, além disso, o GoogLenet possui dois pontos de saída da rede para evitar problemas de *Vanishing Gradient*. Como esse gradiente é retropropagado para as camadas anteriores, a multiplicação que acontece repetidamente entre elas acaba deixando o valor do gradiente quase inexpressivo, resultando em um desempenho saturado ou problema de degradação da rede.

6 ResNet

Possuindo uma arquitetura profunda expressiva, a ResNet [He et al., 2016], surpreende por conseguir ter respostas positivas em relação a performance e saídas desejadas, mesmo contendo muito mais camadas convolucionais em comparação com as demais redes. No entanto, o conjunto de camadas apresentado na ResNet não foi construído

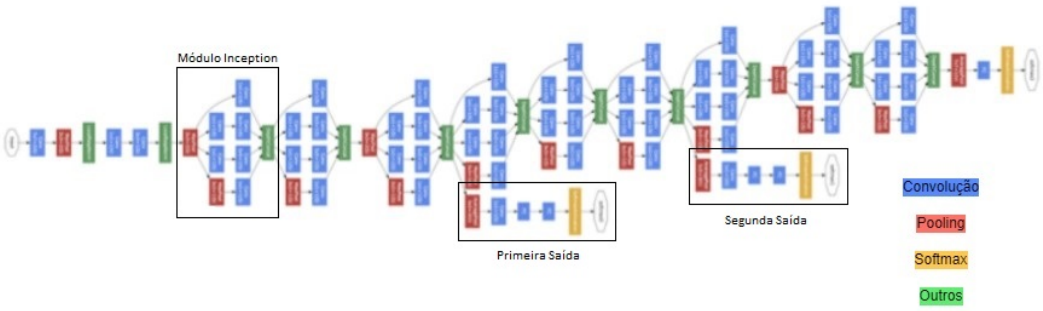

Figura 4: Rede GoogLenet

apenas empilhando-as, já que redes profundas com inúmeras camadas empilhadas sofrem com o problema de *Vanishing Gradient*. A ideia introduzida por He et al. [2016], para resolver os problemas relacionados a profundidade da rede, é chamada de *Identity Shortcut Connection*, onde, nada mais é que pular uma ou mais conexões, ignorando o treinamento de algumas camadas e se conectando diretamente à próxima saída não ignorada. Sendo representado não mais como *F(x)*, e sim como *F(x)+ x* (Fig. 5).

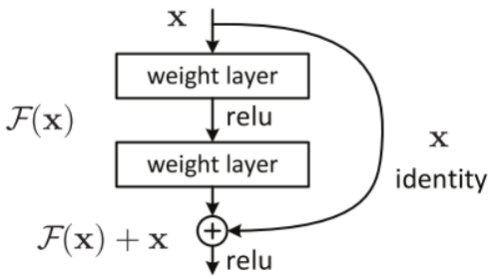

Figura 5: Identity Shortcut Connection

7 Mecanismo de Atenção

Introduzido em 2016, o Mecanismo de Atenção [Bahdanau et al., 2014] é um componente da arquitetura de redes, que foi criado através de uma dificuldade que redes do tipo seq2seq tinha de traduzir longas entradas de texto. Sua função é dar foco em determinados aspectos dentro do processamento de features; gerenciando o que pode ser de fato algo importante para o aprendizado de uma rede. Com essa técnica, a profundidade de rede é aumentada consideravelmente, e naturalmente o poder de processamento acaba sendo maior para a realização do aprendizado, porém, como citado anteriormente, aprendizados utilizando arquiteturas residuais diminui a necessidade de uma grande quantidade de poder computacional. Pouco tempo depois, essa tecnologia foi introduzida no campo de classificação de imagens, sendo uma técnica eficaz no aumento do desempenho e nos resultados positivos ao ser utilizada com arquiteturas de redes residuais. A arquitetura de rede residual utilizando módulos de atenção mostrada nesse projeto, foi projetada e realizada por Wang et al. [2017]. Ela é composta por

três hiperparâmetros: p, t e r, nos quais são responsáveis por quantificar o número de Unidades Residuais de pré-processamento antes de dividir o módulo de Atenção em dois tipos: *mask branch* e *trunk branch*, outro por representar o número de Unidades Residuais no *trunk branch*, e outro denota o número de Unidades Residuais entre as camadas de *pooling*.

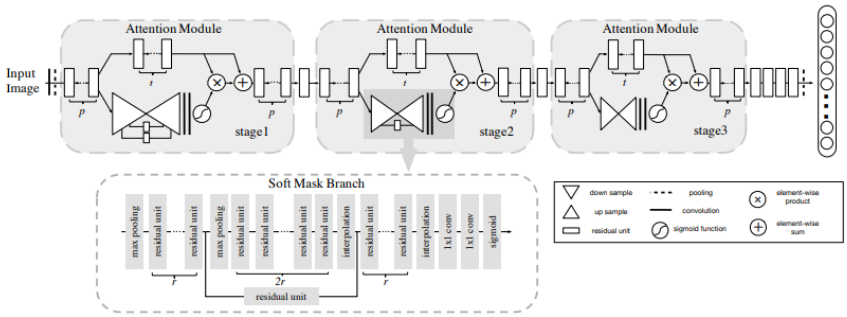

Figura 6: Mecanismo de Atenção

8 Transformers

A arquitetura de Transformers foi criada com o objetivo de lidar com problemas de sequencias ou seq2seq [Vaswani et al., 2017]. A técnica é baseada em modelos de atenção, apresentando uma evolução arrojada do conceito de camadas de atenção, seguidas de camadas simples de feedfoward e operações matriciais. A primeira aplicação de transformers foi construída para tradução de textos em diferentes línguas e apresentavam camadas de encoder e decoder, porém hoje já existem outras aplicações para essa arquitetura com diversas modificações, sendo imagens uma delas Dosovitskiy et al. [2020], Devlin and Chang [2018], Zhu and Luo [2022]

Depois que estudos mostraram que o Vision Transformer (ViT) [Dosovitskiy et al., 2020] obteve um desempenho consideravelmente ao ser usado em algumas bases de dados de imagens, mas ainda sim precisaria ser passado por um pré-treinamento em grande escala para obter resultados semelhantes aos de redes com arquiteturas Convolocionais. Estudiosos perceberam que as camadas das redes Transformers tem certos problemas em relação a vieses indutivos, sendo assim, necessário uma grande quantidade de dados para poder compensar e gerar bons resultados, diferente de arquiteturas Convolucionais, onde possuem esses vieses.

A rede CoAtNet Dai et al. [2021] surgiu a partir desses dois pontos. Com a combinação desses dois estudos foi projetado uma arquitetura que visa a capacidade de generalização sem prejudicar o desempenho da rede. Sendo que as arquiteturas Convolucionais têm melhor generalização, enquanto as camadas de atenção, utilizadas em ViT, obtém melhor capacidade.

Experimentos mostram que redes CoAtNets funcionam muito bem, obtendo resultados impressionantes e de última geração, sendo em conjuntos de dados mais restritivos em relação a *features* ou não, ou até mesmo em diferentes tamanhos de dados. Sendo elas, uma das redes com maiores resultados positivos atualmente.

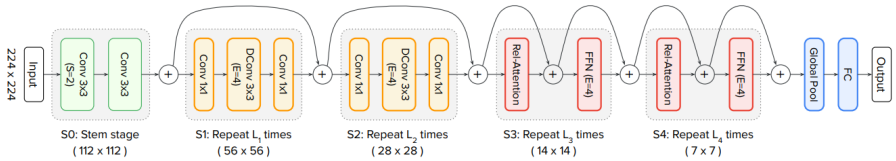

Figura 7: Arquitetura CoAtNet

9 Metodologia

O projeto foi iniciado através da tabela de dados *Normal-depth sample with new debiasing method*, trabalhada por Hart et al. [2016] e encontrada na base de dados Galaxy Zoo [Bamford et al., 2009], onde apresenta características diversas de diferentes galáxias, de forma numérica, incluído seus determinados rótulos (Elíptica, Espiral e suas subclassificações). Em seguida, essa tabela foi correlacionada com uma segunda tabela, *Galaxy Zoo 2: Images from Original Sample*, como citada na sessão 2, na qual contém informações adicionais, que foram utilizadas para encontrar as imagens de cada galáxia descrita na primeira tabela. Com isso, foram obtidas as imagens e seus determinados rótulos.

O conjunto de imagens foi de 239.029 galáxias coloridas, no qual foi separado em duas categorias principais intituladas como espirais (S) e elípticas (E), comparado com as classificações morfológicas por Edwin Hubble [Hubble, 1926], descritas na seção 3 desse trabalho, não foi considerado o rótulo de classificação Irregular (I), por conta da inexistência dessa categoria no conjunto de dados obtidos pelo Galaxy Zoo 2, sendo apenas referenciado como uma subcategoria em uma pequena quantidade, cerca de 5.837 imagens. Também foram unificados os rótulos Espirais barradas (SB) e Espirais ordinais (S) pela alta similaridade das características que ambas possuem, colocando um foco maior da classificação na diferenciação entre Elípticas (E) e Espirais (S+SB). Desse total, 59,2% das imagens foram intituladas como S, enquanto os outros 40,8% foram intituladas como E. Dentro de cada categoria, foi separado amostras para treino, validação e teste, onde as proporções ficaram como 70%, 10% e 20%, respectivamente. O trabalho foi realizado com dois tipos de amostra de imagens: seu estado original, onde nenhum tipo de modificação, além do redimensionamento das imagens, foi feito, e seu estado segmentado, onde foi criado uma segmentação que consiste basicamente de capturar a componente conexa mais próxima ao centro, dessa forma excluindo pontos que são considerados como "sujeiras" ou objetos extras da imagem. Com isso, será possível fazer uma comparação entre os resultados para os dois tipos de amostras de imagens (Fig. 8).

Para as classificações das imagens, foram utilizadas 9 redes neurais através da biblioteca Keras, sendo duas delas, utilizadas em conjunto a Módulos de Atenção, no qual foi descrito na sessão 7. Cada uma das redes exigiu uma determinada dimensão de imagem conforme especificações de suas arquiteturas, como entrada (Tabela 1).

Após a apresentação das imagens para cada uma das redes mencionadas, os resultados foram avaliados utilizando as métricas de desempenho de modelos de classificação, como matriz de confusão, precisão, acurácia, f1-score e recall.

Figura 8: Imagens Originais e Segmentadas

Rede	Tamanho da Entrada
AlexNet	227x227
Attention ResNet 56	224x224
Attention ResNet 92	224x224
CoAtNet0	224x224
GoogLeNet	224x224
Inception ResNet V2	299x299
Inception V3	299x299
MobileNet V2	424x424
ResNet152 V2	224x224

Tabela 1: Dimensões das Imagens

10 Resultados

Como descrita na sessão anterior, a metodologia aplicada nos dois diferentes conjuntos de imagens, denominados como originais e segmentadas, obteve os seguintes resultados (Tabela 2 e Tabela 3).

Grupo de Arquitetura	Rede	Acurácia	VP	VN	FP	FN	Precisão		Recall		F1-Score	
							Elíptica	Espiral	Elíptica	Espiral	Elíptica	Espiral
CNNs Básicas	AlexNet	82,9%	47,8%	35,1%	5,8%	11,4%	75,6%	89,3%	85,9%	80,8%	80,4%	84,8%
	MobileNet V2	83,2%	52,8%	30,4%	10,4%	6,4%	82,7%	83,5%	74,5%	89,3%	78,4%	86,3%
Inception	GoogLeNet X	83,8%	49,7%	34,1%	6,7%	9,5%	78,3%	88,1%	83,6%	84,0%	80,8%	86,0%
	GoogLeNet X1	84,1%	49,8%	34,2%	6,6%	9,3%	78,6%	88,3%	83,8%	84,2%	81,1%	86,2%
	GoogLeNet X2	83,9%	50,0%	34,0%	6,9%	9,2%	78,7%	87,9%	83,1%	84,4%	80,8%	86,1%
	Inception ResNet V2	83,5%	48,6%	34,9%	6,0%	10,5%	76,8%	89,1%	85,4%	82,2%	80,9%	85,5%
	Inception V3	82,8%	46,0%	36,8%	4,1%	13,2%	73,7%	91,9%	90,1%	77,8%	81,0%	84,2%
ResNet	Attention ResNet 56	83,7%	49,9%	33,8%	7,0%	9,3%	78,5%	87,6%	82,8%	84,3%	80,6%	86,0%
	Attention ResNet 92	82,4%	47,6%	34,8%	6,0%	11,6%	75,0%	88,8%	85,3%	80,4%	79,8%	84,4%
	ResNet152 V2	83,6%	50,1%	33,5%	7,3%	9,0%	78,8%	87,2%	82,1%	84,7%	80,4%	86,0%
Transformer	CoAtNet0	84,4%	51,3%	33,1%	7,7%	7,9%	80,8%	86,9%	81,1%	86,7%	81,0%	86,8%

Tabela 2: Resultados Originais

Grupo de Arquitetura	Rede	Acurácia	VP	VN	FP	FN	Precisão		Recall		F1-Score	
							Elíptica	Espiral	Elíptica	Espiral	Elíptica	Espiral
CNNs Básicas	AlexNet	82,7%	51,7%	31,0%	9,9%	7,4%	80,7%	84,0%	75,9%	87,4%	78,2%	85,7%
	MobileNet V2	82,7%	51,5%	31,2%	9,7%	7,6%	80,3%	84,2%	76,4%	87,1%	78,3%	85,6%
Inception	GoogLeNet X	83,9%	49,9%	34,0%	6,8%	9,3%	78,5%	87,9%	83,3%	84,3%	80,8%	86,1%
	GoogLeNet X1	84,0%	49,9%	34,1%	6,7%	9,3%	78,7%	88,1%	83,5%	84,4%	81,0%	86,2%
	GoogLeNet X2	83,8%	50,1%	33,7%	7,2%	9,1%	78,8%	87,5%	82,5%	84,7%	80,6%	86,1%
	Inception ResNet V2	83,9%	51,3%	32,6%	8,2%	7,9%	80,5%	86,1%	79,8%	86,7%	80,2%	86,4%
	Inception V3	83,7%	52,5%	31,2%	9,6%	6,7%	82,4%	84,5%	76,4%	88,7%	79,3%	86,6%
ResNet	Attention ResNet 56	83,4%	50,3%	33,0%	7,8%	8,8%	78,9%	86,6%	80,9%	85,1%	79,9%	85,8%
	Attention ResNet 92	83,3%	50,6%	32,7%	8,2%	8,5%	79,3%	86,1%	80,0%	85,6%	79,6%	85,9%
	ResNet152 V2	82,2%	52,0%	30,2%	10,6%	7,2%	80,9%	83,0%	74,0%	87,9%	77,3%	85,4%
Transformer	CoAtNet0	84,2%	50,1%	34,1%	6,7%	9,1%	79,0%	88,1%	83,5%	84,7%	81,2%	86,4%

Tabela 3: Resultados Segmentadas

Foi notado que entre os diferentes grupos de arquiteturas, no qual foram citados nesse projeto, não houve diferenças significativas. A média de acurácia das redes, tendo como base os resultados de imagens originais e imagens segmentadas, ficou dentro dos 83,7%. A média dos resultados em acurácia das imagens originais ficou em 83,6%, e das imagens segmentadas; em 83,7%.

A pouca diferença entre as imagens segmentadas e as originais indicam que muito provavelmente as redes se adaptaram a considerar apenas as regiões a imagem de maior interesse, descartando assim partes periféricas das imagens. Dessa forma o processo de segmentação utilizado se torna desnecessário e redundante para essas redes.

A rede CoAtNet0, como uma rede do grupo Transformer, obteve o melhor desempenho em acurácia, tanto utilizando as imagens originais, quanto as imagens segmentadas. Ficando com 84,4% e 84,2%, respectivamente.

Os menores resultados ficaram com o grupo de arquitetura ResNet, onde utilizado entradas de imagens originais e módulos de atenção, a rede intitulada como AttentionResNet92, obteve a acurácia de 82,4%, enquanto a rede ResNet 152V2 obteve a acurácia de 82,2% utilizando as imagens segmentadas como entrada. Esse sendo o menor resultado de todo o experimento.

Embora a AlexNet seja um rede mais antiga, os seus resultados não tiveram grande discrepância comparados com as arquiteturas mais atuais, obtendo, até mesmo, melhores resultados em alguns testes comparada com redes que obtiveram os menores resultados. Utilizado imagens originais como entrada, ficou com a acurácia de 82,9%, enquanto com as imagens segmentadas, a acurácia foi para 82,7%.

Dentro do grupo de redes do tipo Inception, o melhor e pior resultado em acurácia ficaram dentro do experimento com imagens originais, onde a saída intermediária do GoogLeNet (X1) ficou com 84,1% e a rede Inception V3; 82,8%. Em relação ao GoogLeNet, sua saída intermediária (X1) também obteve o melhor resultado, utilizando as imagens segmentadas, em comparação com as suas demais saídas, com a acurácia de 84,0%. Tendo a diferença de 0,3% e 0,2% respectivamente, comparado com o pior resultado da rede, que ficou em 83,8%, tanto com entradas de imagens originais, quanto com imagens segmentadas. Já na MobileNet V2, seu melhor resultado ficou em 83,2%, utilizando as imagens originais como entrada, já utilizando imagens segmentadas, sua acurácia foi para 82,7%.

Dessa forma as diferenças sutis de acurácia entre redes mais simples até as mais sofisticadas demonstra que os detalhes pequenos entre as imagens está presente em poucos pixels, e a maioria dessas arquiteturas foram projetas para trabalhar com imagens em que os detalhes necessários para a classificação são maiores e estão espalhados de forma

mais distribuída.

11 Conclusão

É de se observar que a aplicação das diferentes arquiteturas de classificação apresentadas nesse projeto, obtiveram resultados bem próximos; sendo elas, arquiteturas com diferentes características, mas que trabalham para o mesmo objetivo. As redes mais antigas e com a menor complexidade computacional obtiveram resultados próximos as de redes mais profundas e arquiteturas mais complexas, podendo ser notado que, nesse projeto, essas características não foram fundamentais para um melhor resultado, sendo uma possível causa, as pequenas dimensões das imagens e micro diferenças presentes em poucos pixels, diminuindo as vantagens tecnológicas que as arquiteturais mais atuais podem ter em cima de arquiteturas mais antigas ou mais simples.

Referências

James S Dunlop, RJ McLure, T Yamada, M Kajisawa, JA Peacock, Robert G Mann, DH Hughes, Itziar Aretxaga, TWB Muxlow, AMS Richards, et al. Discovery of the galaxy counterpart of hdf 850.1, the brightest submillimetre source in the hubble deep field. *Monthly Notices of the Royal Astronomical Society*, 350(3):769–784, 2004.

Andrew J Bunker, Stephen Wilkins, Richard S Ellis, Daniel P Stark, Silvio Lorenzoni, Kuenley Chiu, Mark Lacy, Matt J Jarvis, and Samantha Hickey. The contribution of high-redshift galaxies to cosmic reionization: new results from deep wfc3 imaging of the hubble ultra deep field. *Monthly Notices of the Royal Astronomical Society*, 409 (2):855–866, 2010.

Linhua Jiang, Nobunari Kashikawa, Shu Wang, Gregory Walth, Luis C Ho, Zheng Cai, Eiichi Egami, Xiaohui Fan, Kei Ito, Yongming Liang, et al. Evidence for gn-z11 as a luminous galaxy at redshift 10.957. *Nature Astronomy*, 5(3):256–261, 2021.

Wen Xiao-Qing and Yang Jin-Meng. Classification of star/galaxy/qso and star spectral types from lamost data release 5 with machine learning approaches. *Chinese Journal of Physics*, 69:303–311, 2021.

Steven P Bamford, Robert C Nichol, Ivan K Baldry, Kate Land, Chris J Lintott, Kevin Schawinski, Anže Slosar, Alexander S Szalay, Daniel Thomas, Mehri Torki, et al. Galaxy zoo: the dependence of morphology and colour on environment. *Monthly Notices of the Royal Astronomical Society*, 393(4):1324–1352, 2009.

Ross E Hart, Steven P Bamford, Kyle W Willett, Karen L Masters, Carolin Cardamone, Chris J Lintott, Robert J Mackay, Robert C Nichol, Christopher K Rosslowe, Brooke D Simmons, et al. Galaxy zoo: comparing the demographics of spiral arm number and a new method for correcting redshift bias. *Monthly Notices of the Royal Astronomical Society*, 461(4):3663–3682, 2016.

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.

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.

Zihang Dai, Hanxiao Liu, Quoc V Le, and Mingxing Tan. Coatnet: Marrying convolution and attention for all data sizes. *Advances in Neural Information Processing Systems*, 34:3965–3977, 2021.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

Paulo H Barchi, RR de Carvalho, Reinaldo R Rosa, RA Sautter, Marcelle Soares-Santos, Bruno AD Marques, Esteban Clua, TS Gonçalves, C de Sá-Freitas, and TC Moura. Machine and deep learning applied to galaxy morphology-a comparative study. *Astronomy and Computing*, 30:100334, 2020.

Edwin P Hubble. Extragalactic nebulae. *The Astrophysical Journal*, 64, 1926.

Chris M Bishop. Neural networks and their applications. *Review of scientific instruments*, 65(6):1803–1832, 1994.

Yann LeCun, Patrick Haffner, Léon Bottou, and Yoshua Bengio. Object recognition with gradient-based learning. In *Shape, contour and grouping in computer vision*, pages 319–345. Springer, 1999.

M I Jordan. Serial order: a parallel distributed processing approach. technical report, june 1985-march 1986. 5 1986. URL https://www.osti.gov/biblio/6910294.

David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

Linda G Shapiro, George C Stockman, et al. *Computer vision*, volume 3. Prentice Hall New Jersey, 2001.

Pascal Getreuer. A survey of gaussian convolution algorithms. *Image Processing On Line*, 2013:286–310, 2013.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

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. doi: 10.1007/s11263-015-0816-y.

Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164, 2017.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017. URL https://arxiv.org/abs/1706.03762.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Jacob Devlin and Ming-Wei Chang. Open sourcing bert: State-of-the-art pre-training for natural language processing. *Google AI Blog*, 2, 2018.

Qihao Zhu and Jianxi Luo. Generative pre-trained transformer for design concept generation: an exploration. *Proceedings of the Design Society*, 2:1825–1834, 2022.