Transfer Learning Applied to a Classification Task: a Case Study in the Footwear Industry

Fernando Gabriel Bloedorn and Carine Geltrudes Webber

***Abstract*—Convolutional Neural Networks are a widely used method for image classification. They are part of the Deep Learning area, whose main advantage is the fact that they do not require an human support to extract features from the images. In the context of the footwear industry, they represent a useful computational resource, being applied for style classification problems, machine vision, among others. This article aims to evaluate the performance of transfer learning methods for the purpose of hierarchical classification of new items of footwear products. For this purpose, a dataset composed by 5,177 images of women's shoes was built. A pretrained architecture was selected to be refined, in order to produce a classification model. As a main result of this study, we confirm that the use of transfer learning speeds up deep neural nets training, allowing outstanding results through a VGG16 architecture. In terms of accuracy, the results achieved 99.97% and 98.42% for classifying respectively footwear categories and subcategories.**

***Index Terms*—Convolutional Neural Networks, Image Classification, Transfer Learning, VGG16.**

I. Introdução

Um dos principais ramos do mundo da moda é o de calçados - em 2020 o mercado global foi US$ 271 trilhões com a projeção de chegar a US$ 328 trilhões em 2027 [1]. Os calçados são produtos de consumo popular, com variedade suficiente no mercado para que vários negócios sejam criados em torno deles. Eles são, em grande parte, comercializados pelo seu apelo visual [2]. Além disso, os calçados são diferenciados principalmente em três características visuais: forma, textura e cor [3]. As Redes Neurais Convolucionais (CNNs) são conhecidas por sua capacidade de aprender formas, texturas e cores básicas, tornando-as aderentes para a tarefa de classificação de imagens de calçados [4].

A classificação de imagens de moda oferece às empresas do ramo uma maneira de entender, categorizar, agrupar, associar e vincular os seus produtos. Ela pode ser usada na categorização por suas marcas, tipos, estilos, etc, assim como para sistemas de recomendação e filtragem de pesquisas em meios digitais. Muitos modelos de redes neurais foram aplicadas com sucesso à classificação de imagens, porém apenas alguns deles podem ser aplicados a imagens de moda, devido à sua natureza. Por exemplo, tais imagens podem pertencer a um tópico de alto nível, tal como: *t-shirt*, saia ou calçado esportivo, mas também a um nível mais específico, tal como: sapato de salto alto e sapato de salto médio. Cada empresa pode ter sua própria hierarquia, o que representa um grande desafio para a tarefa de classificação de seus produtos [5].

Nos últimos anos as CNNs obtiveram grande sucesso em problemas da área de aprendizado de máquina envolvendo visão computacional [6]. Graças à robustez da extração de recursos de uma CNN, os pesquisadores fazem uso dela em uma variedade de aplicações [7]. As CNNs são um método computacional, que integra elementos do processamento de imagem aos princípios das Redes Neurais Perceptron Multicamadas, sendo aplicado a classificação de imagens com elevada acurácia. Porém, treinar os modelos CNN com um grande conjunto de dados é um processo custoso que demanda grande esforço computacional e inúmeros ajustes, para que produzam modelos de classificação acurados. Uma abordagem baseada no aprendizado por transferência (*transfer learning*) que considera que um modelo pré-treinado é retreinado para um novo problema, pode ser a solução para tal cenário [8]. O *transfer learning* ocorre quando um modelo de *machine learning* é usado para tratar um problema de classificação e o mesmo modelo é usado como ponto de partida para tratar um conjunto de dados distinto [9]. Desta forma, obtém-se melhora no desempenho e redução no tempo e esforços gastos na fase de treinamento [10].

Ao longo dos anos, vários modelos de CNN pré-treinados foram propostos na literatura, após obterem um excelente desempenho no conjunto de dados ImageNet [11], como AlexNet, GoogleNet e VGG [2]. Um dos modelos mais importantes é o VGG, que possui variações como VGG11, VGG13, VGG16 e VGG19, com base no número de camadas convolucionais incluídas [12]. Os modelos pré-treinados se aplicam às tarefas de classificação de imagens em grande escala para salvar os pesos dos parâmetros de treinamento finais. Quando o conjunto de dados de treinamento não é grande o suficiente para treinar o modelo CNN de ponta a ponta, uma rede neural pré-treinada é usada como método para economizar recursos. A vantagem em usar-se um modelo pré-treinado é que os recursos aprendidos podem ser transferidos para dados diferentes, ajustando-se de forma eficiente a rede com um conjunto de dados menor. O *fine-tuning* dos parâmetros de modelos CNN pré-treinados com conjuntos de dados menores produz um bom desempenho na verificação do modelo [13].

Neste contexto, o objetivo deste trabalho é avaliar o uso de um processo de *transfer learning* para um problema de classificação de calçados. Por meio de um estudo de caso em uma empresa de calçados femininos, este artigo buscou

Fernando Gabriel Bloedorn, Área de Exatas e Engenharias, Universidade de Caxias do Sul (UCS), Caxias do Sul, Rio Grande do Sul, Brazil, e-mail: fgbloedorn@ucs.br

Carine Geltrudes Webber, Área de Exatas e Engenharias, Universidade de Caxias do Sul (UCS), Caxias do Sul, Rio Grande do Sul, Brazil, e-mail: cgwebber@ucs.br

empregar a arquitetura VGG16 por meio de *transfer learning*.

O restante deste artigo está organizado da seguinte forma. A seção 2 apresenta a revisão sistemática de trabalhos relacionados. A seção 3 mostra os materiais e métodos. A seção 4 oferece uma visão geral do desenvolvimento da proposta. A seção 5 apresenta os resultados do experimento e a seção 6 conclui este artigo.

II. Trabalhos Relacionados

A revisão de literatura é a metodologia de pesquisa mais utilizada quando se deseja mapear trabalhos e soluções relacionadas. Por meio de um processo de revisão sistemática analisou-se artigos e definiu-se o escopo da proposta do presente artigo.

A. Revisão Sistemática

Este artigo utilizou o processo de revisão sistemática cujo objetivo é disponibilizar um resumo das evidências relacionadas a uma estratégia de intervenção específica, mediante a aplicação de métodos explícitos e sistematizados de busca, apreciação crítica e síntese das informações selecionadas. O processo seguido é composto por cinco passos [14]. Cada passo é descrito a seguir:

1) Definindo a pergunta: uma boa revisão sistemática requer uma pergunta ou questão bem formulada e clara. Para o presente trabalho a pergunta a ser respondida é: "a partir do que já foi proposto na literatura, é possível desenvolver um modelo classificador de calçados a partir das Redes Convolucionais?"
2) Buscando a evidência: o início da busca de evidências ocorre a partir da definição de termos ou palavras chave, seguida das estratégias de busca e definição das bases e fontes de dados. Neste artigo definiu-se como termos de busca: "*Convolutional Neural Networks Shoe*", "*Convolucionais Neural Networks Image*", "*Convolucionais Neural Networks Fashion*", "VGG16" "VGG19", "*Transfer Learning*" e "*Fine-tuning*". Como fontes de dados foram utilizadas as seguintes plataformas digitais: "*Science Direct*", "*IEEE Xplore*", "*ACM Digital Library*" e "*Research Gate*".
3) Revisando e selecionando os estudos: faz-se a avaliação dos títulos e resumos (*abstracts*) identificados na busca inicial. Quando o título e o resumo não são esclarecedores deve-se buscar o artigo na íntegra. Além disso, são definidos critérios de inclusão e exclusão. Nesta pesquisa, como critérios de inclusão e exclusão, foram: trabalhos publicados nos últimos cinco anos; que apresentaram maior relevância no ramo da moda; que obtiverem resultados satisfatórios e; que fizeram uso de *transfer learning* e/ou *fine-tuning*.
4) Analisando a qualidade metodológica dos estudos: a qualidade de uma revisão sistemática depende da validade dos estudos incluídos nela. No presente trabalho fez-se uma escala de pontuação baseada na relevância de cada artigo de acordo com o cenário estudado utilizando-se a pontuação de 1 - menos relevante - até 5 - mais relevante.
5) Apresentando os resultados: pode-se destacar em um quadro as características principais dos artigos analisados, como autores, ano de publicação, grupos de comparação e principais resultados. Após a análise dos estudos, os artigos foram ordenados por relevância e selecionou-se os sete mais relevantes que posteriormente foram detalhadamente estudados (Tabela 1).

B. Artigos Relacionados

Realizou-se um estudo aprofundado de cada um dos sete artigos mais relevantes definidos na revisão sistemática. No intitulado "*Building Image-Based Shoe Search Using Convolutional Neural Networks*" [2], os autores exploraram as redes neurais aplicadas à classificação de imagens de calçados e similaridade dos mesmos. Utilizou-se um conjunto de dados de mais de 30.000 imagens, onde buscou-se classificar cada um na sua categoria apropriada e mostrar os cinco mais semelhantes ao conjunto de dados. Com a intenção de fazer uso de *transfer learning* nas duas tarefas os autores fizeram uso do framework para *deep learning Caffe* [15] com três redes neurais convolucionais: ViggyNet Small, ViggyNet Large e VGGNet. As duas primeiras redes são versões menores da VGGNet. Na tarefa de classificação, a execução da ViggyNet Small resultou em uma acuracidade de 92%, enquanto que ViggyNet Large obteve 64%. A rede VGGNet não pode ser utilizada devido a limitações de recursos computacionais. Já para a tarefa da busca de similaridade a VGGNet obteve o melhor resultado com 75,6% de precisão, contra 62,6% da ViggyNet Small e 69,4% da ViggyNet Large. Por fim, os autores indicam limitações no uso da VGGNet com *transfer learning* e *fine-tuning* devido a disponibilidade computacional.

Várias arquiteturas CNN foram propostas para melhorar o desempenho nas tarefas de classificação de imagens, entre elas, AlexNet, VGG16 e VGG19. No trabalho de Shaha e Pawar [7], foi utilizado *transfer learning* e *fine-tuning* da rede VGG19 em comparação com as redes AlexNet e VGG16 para a classificação de imagens. Na fase de treinamento utilizaram a base de dados ILSVRC [11], que consiste em 22.000 categorias de objetos. Através desta base, realizaram o *transfer learning* e o *fine-tuning* da rede, obtendo o modelo da VGG19 treinado. Já na fase de testes, realizaram dois experimentos, cada um com um conjunto de dados diferentes: CalTech256 - conjunto de 256 classes com pelo menos 80 imagens em cada [16] - e GHIM10K - conjunto de 10.000 imagens e 20 classes [17]. Em ambas avaliações a rede VGG19 obteve melhores resultados. Os autores citam que a VGG19 foi composta por algumas unidades de RELU convolucionais extras no meio da rede em comparação a VGG16 e, com essa mudança, a arquitetura obteve melhores resultados para a tarefa de reconhecimento de objetos.

O trabalho de Khanuja [18] visou usar um conjunto de dados de 17.500 imagens de produtos pertencentes a cinco categorias de uma plataforma de comércio eletrônico e desenvolver um algoritmo para classificar com precisão os produtos em suas respectivas categorias, levando o menos tempo possível. O objetivo foi testar o desempenho de *transfer learning* em relação às redes convolucionais tradicionais. Para o modelo

TABELA I
REVISÃO SISTEMÁTICA DOS TRABALHOS RELACIONADOS.

Título	Autor(es) e ano	Dataset	Ferramentas	Métricas e Resultados	Relevância
Building Image-Based Shoe Search Using Convolutional Neural Networks	Neal Khosla e Vignesh Venkataraman, 2015	Zappos.com	VGGnet ViggyNet Small ViggyNet Large Transfer learning	ViggyNet Small: 62,6% precisão ViggyNet Large: 69,4% precisão	4
Transfer learning for image classification	Manali Shaha e Meenakshi Pawar, 2018	GHIM10K e Cal-Tech256	AlexNet, VGG16, VGG19. Transfer learning, Fine-tuning	Alexnet: 96,56%/87,31% VGG16: 98,23%/88,24% VGG19: 99,23%/88,88% precisão	5
Optimizing E-Commerce Product Classification Using Transfer Learning	Rashmeet Kaur Khanuja, 2019	E-commerce de produtos diversos	VGG16 Transfer learning Freezing Fine-tuning	VGG16: 73% acurácia	5
E-Commerce Product Image Classification using Transfer Learning	Bineet Jha, Sivasankari G. G e Venugopal Krishnappa, 2021	Fashion-MNIST	VGG19 InceptionV3 Transfer learning	VGG19: 88% acurácia Inception V3: 78% acurácia	5
Fashion and Apparel Classification using Convolutional Neural Networks	Alexander Schindler e Thomas Lidy, 2018	Asos-EU, Farfetch e Zalando	VGG16 VGG19 InceptionV3 Transfer learning Fine-tuning	VGG19: 88% acurácia VGG16: 71% acurácia Inception V3: 85% acurácia	5
Hierarchical convolutional neural networks for fashion image classification	Yian Seo e Kyung-shik Shin, 2019	Fashion-MNIST	VGG16 VGG19 Fine-tuning	VGG16: 93,52% acurácia VGG19: 93,33% acurácia	5
Condition-CNN: A hierarchical multi-label fashion image classification model	Brendan Kolisnik, Isaac Hogan e Farhana Zulkernine, 2021	Kaggle Fashion Product Images	VGG16 Transfer learning	Acurácia: 99,8% level 1 98,1% level 2 91% level 3	4

com *transfer learning*, a VGG16 foi treinada no conjunto de dados ImageNet. Ao se comparar os resultados com um conjunto de dados de 3.039 imagens, observou-se que o tempo gasto foi de três horas para atingir 79% de precisão usando uma CNN tradicional, enquanto que no modelo com *transfer learning* o resultado foi de 85% em 16,96 minutos. Por fim, o autor sugere que, como trabalho futuro, pode-se fazer uso da solução apresentada para a classificação em níveis mais profundos, como as subcategorias das categorias.

O *transfer learning* pode ser usado para reduzir o tempo de treinamento e validação de um modelo de visão computacional [8]. Jha, Sivasankari e Venugopal [8] propuseram uma abordagem de *transfer learning* para auxiliar na classificação de imagens e busca de similaridade em um site de comércio eletrônico. Para a tarefa de classificação foi selecionada a VGG19 pré-treinada. O *dataset* utilizado foi o Fashion-MNIST, o qual possui 70.000 imagens com 10 classes que foram divididas em 85% para treinamento e 15% para teste. O resultado foi considerado satisfatório pelos autores, chegando ao resultado de 98% de precisão.

No mesmo sentido, Schindler et. al. [19] analisaram cinco arquiteturas aplicadas à classificação de imagens de comércio eletrônico da moda, utilizando três diferentes conjuntos de dados fornecidos por empresas do ramo. As redes analisadas foram: VGG16, VGG19, InceptionV3, Custom CNN e VGG-like. Três tarefas foram testadas nesta proposta: classificação binária da imagem em pessoa ou produto, classificação das categorias das imagens dos produtos e predição de gênero. Na tarefa da classificação dos produtos, foram utilizados dois subconjuntos de dados: um de menor escala contendo 23.305 imagens e outro contendo 234.408, os quais foram aplicados às cinco arquiteturas CNNs. Para a tarefa de classificação abordou-se o treinamento de duas formas: a partir do zero e pré-treinados com *fine-tuning*. Os resultados indicaram que, apesar da grande quantidade e alta qualidade das imagens de moda fornecidas, os modelos pré-treinados e ajustados superam aqueles que foram treinados do zero.

A base de dados Fashion-MNIST também foi utilizada no trabalho "*Hierarchical convolutional neural networks for fashion image classification*" [20], onde se propôs aplicar uma classificação hierárquica para a classificação das imagens da base com o uso de CNNs. Neste trabalho, o *dataset* foi separado de forma hierárquica em três níveis - o primeiro nível com duas categorias, o segundo nível com seis e o terceiro com dez - e, assim, cada imagem processada recebeu três rótulos. As redes utilizadas foram a VGG16 e VGG19 e os autores batizaram o modelo proposto de H-CNN. Os resultados finais foram a acurácia de 93,52% para a VGG16 e 93,33% para a VGG19. Por fim, os autores lembram que para melhorar o desempenho do modelo proposto, e reduzir o tempo gasto no treinamento do classificador, pode-se pré-treinar com o conjunto de dados ImageNet antes de aplicar o modelo, possibilitando a construção de modelos a partir de pequenos conjuntos de dados.

Na classificação hierárquica de imagens, o objeto pode receber vários rótulos definidos em uma hierarquia [21]. No trabalho de Kolisnik, Hogan e Zulkernine [21] foi proposta a implementação de um novo modelo utilizando a arquitetura VGG16. Os autores denominaram esta implementação de Condition-CNN, destinada a classificação hierárquica de imagens. O conjunto de dados utilizado foi o *Kaggle Fashion Product Images*, composto por 41.027 imagens com três níveis de hierarquia de classes (categoria principal com quatro classes, subcategoria com 21 classes e tipo com 45 classes). Os resultados demonstraram uma precisão de 99,8%, 98,1% e 91,0%, respectivamente.

C. Pontos de Destaque dos Artigos Selecionados

Diante do que foi observado nos trabalhos relacionados, o uso de *transfer learning* mostrou-se bastante eficiente para a tarefa de classificação de imagens de moda. As redes da família das VGGs foram usadas em 100% dos artigos analisados. Entre as VGGs, as com mais destaque foram a VGG16 e VGG19 com a base Imagenet. Elas apresentaram excelentes resultados, sendo que para a classificação de forma hierárquica a VGG16 mostrou ser a mais eficiente. A fig. 1 sintetiza a relação das redes utilizadas nos trabalhos.

Desta forma, um caminho promissor parece estar relacionado ao uso de uma rede VGG16 pré-treinada, visando a classificação hierárquica de produtos. Como método de amostragem pode-se empregar o *cross-validation k-fold*. Como métricas nota-se a incidência da acurácia e precisão.

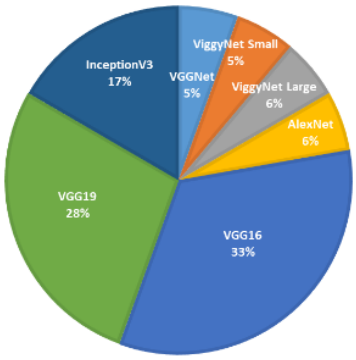

Fig. 1. Redes pré-treinadas identificadas nos trabalhos analisados.

III. Materiais e Métodos

Nesta seção são descritos os materiais e métodos utilizados no presente trabalho.

A. Descrição do Estudo de Caso

Este estudo será aplicado ao cenário de uma empresa de calçados de moda feminina. No seu portfólio a empresa possui uma grande quantidade de calçados em diferentes categorias e subcategorias. De forma geral, a companhia atua em três diferentes ramos: clientes varejistas em todo o Brasil e mais de 30 países em cinco continentes; lojas próprias; e e-commerce próprio. Buscando atender as demandas na rápida dinâmica das novidades da moda, quinzenalmente são criados e lançados novos produtos. Diante deste cenário, faz-se necessário que as informações relacionadas a cada produto estejam rapidamente disponíveis nos seus sistemas para uso dos times nos diferentes ramos do negócio. Além disso, é necessário que as informações estejam cadastradas de forma correta e detalhada.

Os principais objetivos para a empresa na tarefa de classificação de imagens de seus produtos são: aumentar a velocidade na disponibilidade das informações para uso dos times nos diferentes ramos de negócio; reduzir os custos operacionais; enriquecer os metadados e; melhorar o gerenciamento dos produtos.

B. Descrição do Método

O presente trabalho constitui uma pesquisa de natureza exploratória, a qual visa investigar, compreender e aplicar técnicas de *deep learning* no contexto da visão computacional para reconhecimento de imagens de calçados femininos, com o intuito de construir modelos de classificação baseados nas estruturas de redes neurais pré-treinadas da família das VGGs.

O processo de extração de conhecimento é interativo e iterativo, envolvendo várias etapas e decisões a serem tomadas [22]. Seguindo tal processo, esse trabalho foi planejado em seis etapas (Fig. 2): aquisição de imagens [23], pré-processamento das imagens, aplicação de algoritmos, uso de *transfer learning*, comparação dos modelos e seleção do melhor modelo de classificação.

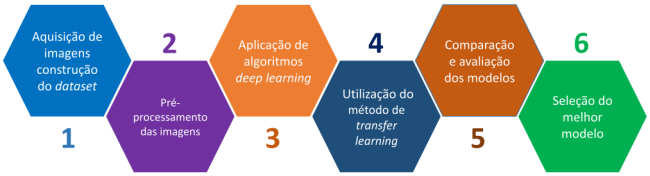

Fig. 2. Etapas do desenvolvimento do trabalho.

(a) Categoria: Tênis. Subcategoria: Sapatilha.

(b) Categoria: Bota. Subcategoria: Coturno.

Fig. 3. Exemplo de imagens dos calçados.

C. Materiais

A plataforma experimental constituída para este estudo compreende um computador com sistema operacional Linux Ubuntu 18.04, contendo uma CPU Intel Xeon Platinum de 8 núcleos e 32 GB de RAM. As ferramentas empregadas são Visual Studio Code 1.62.3 para plataforma de codificação, Python versão 3.6.9 como a linguagem de programação, e as bibliotecas Keras versão 2.6 e Tensorflow versão 2.6.2 [24].

IV. Desenvolvimento

Nesta seção são apresentadas as etapas do desenvolvimento.

A. Etapa 1 - Aquisição de Imagens

Para o processo de aquisição das imagens foram selecionadas as imagens dos produtos criados pela empresa nos últimos cinco anos. As imagens foram coletadas dos servidores da empresa e salvas em uma estrutura de pastas de acordo com sua categoria e subcategoria. Inicialmente, o conjunto de dados resultou em quatro categorias e quarenta e quatro subcategorias. A fig. 3 ilustra exemplos de imagens.

Notou-se que parte das subcategorias continha poucas amostras. Optou-se por eliminar aquelas com menos de 100 instâncias. Desta forma, o conjunto final de imagens se manteve com quatro categorias e dezoito subcategorias, totalizando 5.177 instâncias.

O conjunto de dados foi separado em 60% para treino, 15% para validação e 25% para teste. Para cada um dos conjuntos foi criado um arquivo no formato CSV com as informações do ID, categoria, subcategoria e caminho da imagem. A separação dos dados foi feita de forma estratificada, podendo ser visualizada na fig. 4.

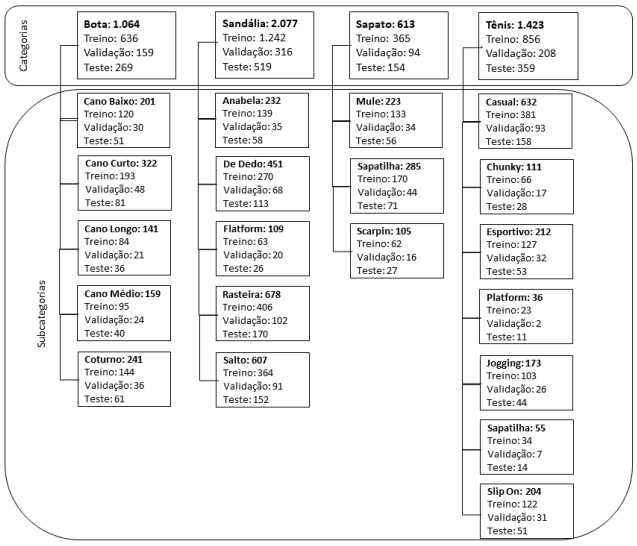

Fig. 4. Separação das instâncias do *dataset*

Fig. 5. Exemplo de imagens criadas após o pré-processamento.

B. Etapa 2 - Pré-processamento de Imagens

Efetuou-se uma série de transformações para pré-processar as imagens. Aplicou-se o redimensionamento para que os valores RGB ficassem entre 0 e 1. As imagens foram rotacionadas aleatoriamente na horizontal em uma escala de até sete graus, ampliadas em até 10% e cortadas em até 10% para adicionar ruído aleatório e assim melhorar a precisão da classificação. Não fez-se necessário redimensionar as imagens pois elas já possuíam a dimensão de 150x200. A fig. 5 mostra alguns exemplos de imagens criadas após a etapa de pré-processamento.

C. Etapa 3 - Aplicação do Algoritmo

Nesta etapa, fez-se uso da arquitetura proposta por Kolisnik, Hogan e Zulkernine [21]. Os blocos da rede original foram alterados para atender as necessidades do presente trabalho. A arquitetura do modelo foi dividida em seis blocos. Os quatro primeiros blocos usam a arquitetura de uma rede VGG16, para a tarefa de extração de características. Já os blocos cinco e seis correspondem às camadas densas de aprendizado e classificação da categoria e subcategoria, conforme observa-se na fig. 6. A rede aprende as associações entre as subcategorias e as categorias, no contexto hierárquico, modelando as probabilidades condicionais como pesos. Empregou-se a arquitetura descrita em dois cenários: com e sem *transfer learning*.

D. Etapa 4 - Transfer Learning

No cenário que utilizou *transfer learning*, partiu-se de um modelo VGG16 pré-treinado com o *dataset ImageNet*. Neste contexto, o método de *transfer learning* permite que as

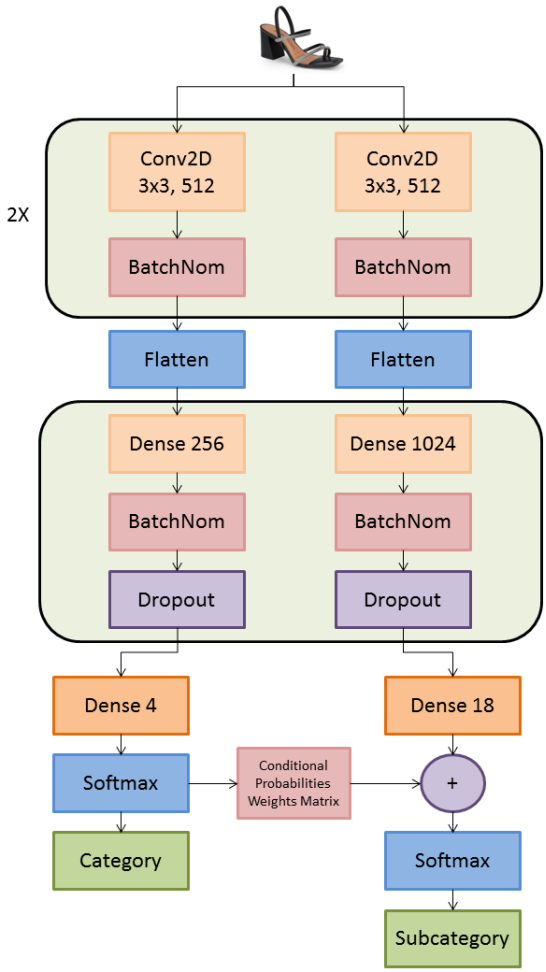

Fig. 6. Arquitetura da Rede VGG16 modificada.

camadas de extração de características, previamente treinadas, sejam transferidas para o novo modelo. O *dataset ImageNet* contém imagens variadas, visualmente distintas das imagens de calçados. Contudo, diversos estudos prévios revelam que as características extraídas nas camadas convolucionais dos modelos pré-treinados com o *dataset ImageNet* tem possibilitado a construção de modelos em variados domínios [6], [9], [25]. De forma similar, busca-se avaliar se as camadas pré-treinadas podem ser aplicadas na classificação de calçados.

E. Etapa 5 - Comparação dos Modelos

Na etapa de comparação dos modelos, as métricas de acurácia, precisão, *recall* e *F1-score* foram relacionadas de acordo com os resultados obtidos na aplicação de cada execução do algoritmo, com os diferentes hiperparâmetros, nas fases de treino e teste. Além disso, para auxiliar no entendimento, foram gerados gráficos de acurácia e perda em cada uma das execuções.

F. Etapa 6 - Seleção do Melhor Modelo

Para fins de testes e seleção do melhor modelo utilizou-se a técnica de otimização de parâmetros denominada *grid search*, na qual cada combinação de hiperparâmetros é testada. O espaço de busca dos hiperparâmetros compreendeu o otimizador

TABELA II
RESULTADOS.

		Validação	Testes
Teste inicial, sem transfer learning 20 épocas, e SGB	Categoria	95,05%	94,61%
	Subcategoria	68,33%	65,31%
Teste final, com transfer learning 50 épocas, e SGB	Categoria	99,87%	95,84%
	Subcategoria	98,42%	76,75%

(Adam, RMSProp, SGD), a taxa de aprendizagem (1e-3, 1e-4, 1e-5), a função de ativação (Relu, Elu, Tanh), o número de neurônios das camadas intermediárias (64, 128, 256, 512, 1024) e constante *momentum* (0.9, 0.95, 1). Realizou-se treinos de 20, 25 e 50 épocas. De acordo com os resultados obtidos, escolheu-se os melhores modelos com e sem *tranfer learning*. Verificou-se que os modelos que utilizaram *transfer learning* apresentaram melhores resultados, como é detalhado na próxima seção.

V. RESULTADOS

O modelo foi avaliado quanto a hierarquia do produto, para definição da sua categoria e subcategoria. O desempenho do modelo é melhor ao prever uma categoria do produto do que subcategoria. Na pesquisa proposta, utilizou-se as métricas de acurácia, precisão, *recall* e *F1-score* para avaliar o modelo de classificação em termos de precisão de predição. Os melhores hiperparâmetros identificados consideram o otimizador como o SGD, a taxa de aprendizagem igual a 1e-3, a função de ativação das camadas intermediárias sendo Relu e nas camadas de saída sendo Softmax, número de neurônios das camadas intermediárias 256 e 1024 e a constante *momentum* igual a 0.9.

Na aplicação da rede inicial os resultados de acurácia foram de 95,05% para a categoria e 68,33% para a subcategoria na fase de treino e 94,61% e 65,31% na fase de testes. Já no resultado final, fazendo uso de *transfer learning* e ajuste de parâmetros, os resultados foram de 99,87% e 98,42% no treino e 95,84% e 76,75% na fase de testes (Tabela II).

Notou-se avanço nos resultados ao fazer-se uso de *transfer learning*. A acurácia do treinamento e da validação melhorou com o aumento do número de épocas, enquanto a perda de treinamento e validação é significativamente reduzida com o aumento das épocas. Verificou-se maior aumento nos resultados na fase de treinamento, enquanto que na fase de testes o aumento do desempenho foi menos significativo.

Entre as categorias, a que teve melhor performance na fase de testes foi a de Sapatos com 97,69% de acertos, enquanto que a de Sandálias, como 95,31%, foi a que apresentou pior resultado.

Já para as subcategorias, as que tiveram os melhores desempenho foram Mule (97,73%), Cano Longo (89,74%), Coturno (88,68%) e Sapatilha (87,50%). Três subcategorias tiveram o resultado de acurácia abaixo de 55%: Flatform (47,73%), Cano Médio (52,63%) e Slip On (54,17%).

VI. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

Este artigo descreveu o desenvolvimento de modelos para a classificação hierárquica de categorias e subcategorias de calçados. Considerou-se o desenvolvimento de modelos com e sem *transfer learning*. Para a classificação hierárquica os melhores resultados foram observados com o uso de *transfer learning*, obtendo-se 99,87% de acurácia para as categorias e 98,42% para as subcategorias. Este resultado foi considerado satisfatório em comparação ao apresentado no trabalho de Kolisnik, Hogan e Zulkernine [21], no qual os resultados apresentados foram de 99,8% e 98,1%, respectivamente.

Verificou-se que com o uso de *transfer learning* os resultados do treinamento obtiveram imediata e considerável melhora em comparação a implementação sem *transfer learning*. Considera-se, portanto, a escolha correta utilizá-lo. A implementação realizada se limitou à arquitetura da VGG16 modificada, proposta no trabalho de Kolisnik, Hogan e Zulkernine [21]. Ao analisar-se os trabalhos relacionados, observa-se que a família VGG se destaca. Em especial, os trabalhos relacionados produziram resultados diversos com as variantes VGGNet ([2]), VGG16 ([7], [18]–[21]) e VGG19 ([7], [8], [19], [20]). Embora os trabalhos utilizem *datasets* diferentes entre si, é possível observar-se que as arquiteturas VGG produzem os melhores resultados. Em grande parte dos casos, não parece ser significativa a diferença dos resultados entre os modelos pré-treinados com VGG16 e VGG19. Comparativamente, o resultado obtido neste trabalho está alinhado aos melhores resultados obtidos pelos trabalhos relacionados (acurácias acima de 98%) [21], [25].

Com relação aos resultados na classificação das categorias, verificou-se as com menos instâncias no conjunto de testes foram as que obtiveram melhores resultados. A categoria sapato, que atingiu o melhor resultado com 97,69% de acurácia, representava 10,01% do conjunto, enquanto que sandália, com o pior resultado, 95,31%, representava 41,03%. Desta forma conclui-se que não houve uma relação entre o número o maior de instâncias e o maior percentual de acertos.

Já para as subcategorias, três delas apresentaram resultados abaixo do esperado com menos de 55% de acurácia. A subcategoria *Flatform*, com 47,73%, foi a que obteve menor percentual de acerto. Esta estava contida dentro de duas categorias: sandália e tênis. Porém, a subcategoria sapatilha, também possuía essa característica, estando contida nas categorias sapato e tênis, e atingiu um resultado de 87,5% de acertos. Dessa forma não pode-se afirmar que, fazer parte de duas categorias, tenha influenciado nos resultados. Ainda, assim como nas categorias, não identificou-se relação entre o número de instâncias no conjunto de testes e os resultados atingidos.

Observou-se que algoritmo empregado se adaptou muito bem ao conjunto de dados da área calçadista, constituindo assim, uma ferramenta apropriada de classificação hierárquica de imagens de calçados para a empresa. Atingiu-se os objetivos previstos de acelerar a tarefa de categorização dos seus produtos, enriquecendo a base de dados dos seus cadastros e melhorando o gerenciamento dos seus produtos.

A necessidade de maior poder computacional para a execução da classificação de imagens, principalmente na fase de treinamento, mostrou-se algo necessário durante a elaboração deste estudo. Foi preciso a aquisição de uma CPU de 8 núcleos e 32 GB de RAM para que atendesse a execução do algoritmo

em tempo satisfatório.

Como trabalhos futuros, pretende-se aumentar o número de imagens do conjunto de dados a fim de melhorar os resultados na classificação, principalmente das subcategorias. Outro aspecto importante será aprofundar os estudo sobre *fine-tuning*, aplicando-o ao algoritmo e, assim, buscando melhores resultados. Além disso, pretende-se utilizar outras arquiteturas, tais como a VGG19 e a InceptionV3, comparando os resultados com os encontrados neste estudo. Por fim, estima-se criar um sistema de gestão dos resultados obtidos na classificação com uma interface onde se possa realizar inspeção visual e curadoria para as imagens com menor percentual de acerto.

REFERENCES

[1] A. More, "Global footwear market and foot care products market size 2021 by trends evaluation, consumer-demand, consumption, recent developments, strategies, leading players updates, market impact and forecast till 2027." [Online]. Available at https://www.globenewswire.com, jul. 2021.

[2] N. Khosla and V. Venkataraman, "Building image-based shoe search using convolutional neural networks," *CS231n course project reports*, pp. 1–7, 2015.

[3] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin, "The qbic project: Querying images by content, using color, texture, and shape.," *SPIE Conference on Storage and Retrieval for Image and Video Databases*, vol. 1908, pp. 173–187, 01 1993.

[4] A. Karpathy, "Visualizing what convnets learn." *CS231n: Deep Learning for Computer Vision,* Stanford University, 2015. [Online]. Available at https://cs231n.github.io/understanding-cnn/.

[5] A. Iliukovich-Strakovskaia, "Using pre-trained models for fine-grained image classification in fashion field." pp. 1-5, 2016. [Online]. Available at https://kddfashion2016.mybluemix.net/.

[6] E. Malvacio and J. C. Duarte, "An assessment of the effectiveness of pretrained neural networks for malware detection," *IEEE Latin America Transactions*, vol. 100, Oct. 2022.

[7] M. Shaha and M. Pawar, "Transfer learning for image classification," in *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 656–660, 2018.

[8] B. K. Jha, S. G. G, and V. K. R, "E-commerce product image classification using transfer learning," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 904–912, IEEE, 2021.

[9] J. Díaz-Ramírez, F. Alvarez-Alvarez, and X. Badilla-Torrico, "Fine-grained geometric shapes: A deep classification task," *IEEE Latin America Transactions*, vol. 20, p. 1051–1057, Jun. 2022.

[10] D. Misal, "Significance of transfer learning in the world of deep learning." (Dec. 17, 2018). Acessed: Nov. 20, 2021. [Online]. Available at https://analyticsindiamag.com/transfer-learning-deep-learning-significance.

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge.," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[12] T. Choudhary, V. Mishra, A. Goswami, and J. Sarangapani, "A transfer learning with structured filter pruning approach for improved breast cancer classification on point-of-care devices," *Computers in Biology and Medicine*, vol. 134, p. 104432, 2021.

[13] A. J. Trappey, C. V. Trappey, and S. Shih, "An intelligent content-based image retrieval methodology using transfer learning for digital ip protection," *Advanced Engineering Informatics*, vol. 48, p. 101291, 2021.

[14] R. Sampaio and M. Mancini, "Estudos de revisão sistemática: Um guia para síntese criteriosa da evidência científica," *Brazilian Journal of Physical Therapy*, vol. 11, no. 1, pp. 83–89, 2007.

[15] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM International Conference on Multimedia*, MM '14, (New York, NY, USA), p. 675–678, Association for Computing Machinery, 2014.

[16] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset." (California Institute of Technology), pp. 1-20, 2007. [Online]. Available at https://authors.library.caltech.edu/7694/.

[17] J. Li and J. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075–1088, 2003.

[18] R. K. Khanuja, "Optimizing e-commerce product classification using transfer learning." Master's Project, Computer Science, San Jose University, San Jose, CA, USA, 2019.

[19] A. Schindler, T. Lidy, S. Karner, and M. Hecker, "Fashion and apparel classification using convolutional neural networks." Proceedings of the 10th Forum Media Technology and 3rd All Around Audio Symposium, St. Poelten, Austria, Nov. 2017, pp. 29-30.

[20] Y. Seo and K. shik Shin, "Hierarchical convolutional neural networks for fashion image classification," *Expert Systems with Applications*, vol. 116, pp. 328–339, 2019.

[21] B. Kolisnik, I. Hogan, and F. Zulkernine, "Condition-cnn: A hierarchical multi-label fashion image classification model," *Expert Systems with Applications*, vol. 182, p. 115195, 2021.

[22] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, eds., *Advances in Knowledge Discovery and Data Mining*. USA: American Association for Artificial Intelligence, 1996.

[23] https://github.com/fernandobloedorn/ImageClassificationFootwear.

[24] *Keras: the Python deep learning API*. (2021). Accessed: Fev 12, 2022. [Online]. Available: https://keras.io/.

[25] M. Shaha and M. Pawar, "Transfer learning for image classification," *Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology*, 2014.

Fernando Gabriel Bloedorn Graduado em Sistemas de Informação pela Universidade Feevale, Pós-graduado em Administração de TI pela Universidade do Vale do Rio do Sinos e Pós-graduado em Ciência de Dados pela Universidade de Caxias do Sul. Atua como Arquiteto de Software, Desenvolvedor de Software, Líder Técnico e Consultor em Tenlogia da Informação. Entusiasta na área de Ciência de Dados, Machine Learning e Inteligência Artificial.

Carine Geltrudes Webber Doutora em Ciência da Computação pela École Doctorale Mathématiques et Informatiques, da Université de Grenoble I Joseph Fourier, França, Mestre (UFRGS) e Graduada (UCS) em Ciência da Computação. Atua como Professor Titular na Área de Conhecimento de Exatas e Engenharias da UCS. Integra o Programa de Pós-graduação em Ensino de Ciências e Matemática, desenvolvendo a linha de pesquisa de IA aplicada ao Ensino. Atua em diversos cursos e projetos de pesquisa nas áreas de Ciência de Dados e Indústria 4.0.