CLASSIFICAÇÃO DE IMAGENS IMOBILIÁRIAS BASEADA EM REDE NEURAL CONVOLUCIONAL EFFICIENTNET V2M

Amanda Rodrigues de Melo
Universidade Federal de Juiz de Fora
Rua José Lourenço Kelmer, s/n - São Pedro, Juiz de Fora - MG, 36036-900
amanda.melo@engenharia.ufjf.br

Leonardo Lima Macedo Pires
Vitrine Urbana Tecnologia Ltda.
Rua Francisco Vaz de Magalhães, 301, Cascatinha, Juiz de Fora - MG, 36.033-340
leopires@semsenha.com

Márcio Lopes de Assis
Vitrine Urbana Tecnologia Ltda.
Rua Francisco Vaz de Magalhães, 301, Cascatinha, Juiz de Fora - MG, 36.033-340
marcio@goldenpar.com

Ricardo Antonio Firjam Filho
Vitrine Urbana Tecnologia Ltda.
Rua Francisco Vaz de Magalhães, 301, Cascatinha, Juiz de Fora - MG, 36.033-340
firjamrick@gmail.com

Kemper Alves de Castro Carlos
Vitrine Urbana Tecnologia Ltda.
Rua Francisco Vaz de Magalhães, 301, Cascatinha, Juiz de Fora - MG, 36.033-340
kemper@topresult.com.br

Eduardo Pestana de Aguiar
Universidade Federal de Juiz de Fora, Vitrine Urbana Tecnologia Ltda.
Rua José Lourenço Kelmer, s/n - São Pedro, Juiz de Fora - MG, 36036-900,
Rua Francisco Vaz de Magalhães, 301, Cascatinha, Juiz de Fora - MG, 36.033-340
eduardo.aguiar@engenharia.ufjf.br

RESUMO

Imagens são capazes de descrever um ambiente e atrair mais atenção em um anúncio de imóveis. Através das imagens as pessoas podem reconhecer diversos atributos de um imóvel, como o tipo de piso, móveis, cor de paredes e até mesmo a vizinhança. Tem se tornado cada vez mais comum a migração do setor imobiliário para o meio digital e com isso a aplicação de inteligência computacional no setor vem crescendo. O presente trabalho visa a implementação do algoritmo de rede neural convolucional EfficientNetV2M para classificação de imagens imobiliárias em seis classes: banheiro, cozinha, fachada, quarto, quintal e sala de estar. Os resultados mostraram resultado de acurácia superiores e resultados de perdas menores a outros modelos de estado-da-arte de redes neurais convolucionais.

PALAVRAS CHAVE. EfficientNetV2M. Classificação. Redes Neurais Convolucionais.

https://proceedings.science/p/157454?lang=pt-br

Tópicos: 11 - IC - Inteligência Computacional, 5- BDA - Big Data e Analytics.

ABSTRACT

Images can describe an environment and attract more attention in a real estate ad. Through the pictures, people can recognize different attributes of a property, such as the type of floor, furniture, wall color and even the neighborhood. Therefore, the migration of the real estate sector to the digital environment has become increasingly common, and the application of computational intelligence in the sector has been growing. The present work aims to implement the EfficientNetV2M convolutional neural network algorithm to classify real estate images into six classes: bathroom, kitchen, frontyard, bedroom, backyard and living room. The results showed higher accuracy results and lower loss results than other state-of-the-art models of convolutional neural networks.

KEYWORDS. EfficientNetV2M. Classification. Convolutional Neural Networks.

Paper topics: 11 - IC - Computer Intelligence, 5- BDA - Big Data and Analytics.

https://proceedings.science/p/157454?lang=pt-br

1. Introdução

O modelo de aprendizado profundo (do inglês *deep learning*) consiste em uma coleção de camadas de processamento que podem aprender vários recursos de dados por meio de vários níveis de abstração [Indolia et al., 2018] [LeCun et al., 2015]. As Redes Neurais Convolucionais (CNNs), também chamada de ConvNet, obtiveram resultados inovadores na última década em uma variedade de campos relacionados ao reconhecimento de padrões; do processamento de imagem ao reconhecimento de voz [Albawi et al., 2017]. As CNNs remontam ao estudo do córtex visual do cérebro e têm sido usadas no reconhecimento de imagens desde a década de 1980 [Géron, 2021]. Cada neurônio responde a estímulos apenas em áreas restritas de campo visual, que é definido como campo receptivo [Rosa et al., 2020]. Os campos receptivos de diferentes neurônios podem se sobrepor e, juntos, revestir todo o campo visual [Géron, 2021]. [Indolia et al., 2018] enumera algumas das principais razões pelas quais a CNN é considerada superior a outros modelos clássicos. Em primeiro lugar, a CNN utiliza o conceito de compartilhamento de peso, onde o número de parâmetros que necessitam de treinamento é substancialmente reduzido, resultando em uma melhor generalização [Arel et al., 2010], então devido a parâmetros menores, a CNN pode ser treinada sem sofrer sobreajuste (do inglês *overfitting*) [Smirnov et al., 2014]. Em segundo lugar, a etapa de classificação é incorporada à etapa de extração de atributos [LeCun et al., 1998]. Por fim, a dificuldade de implementar grandes redes usando modelos gerais de rede neural artificial (RNA) é maior quando comparada a implementação em CNN [Tivive e Bouzerdoum, 2005].

Métodos de inteligencia computacional aplicados em imagens imobiliárias vem sendo amplamente utilizados. [Bappy et al., 2017] propuseram um framework de classificação de imagens de alta confiança, cujas entradas são imagens e as saídas são rótulos. Para isto utilizaram a arquitetura de rede neural recorrente (RNN) memória de curto prazo (do inglês *long short term memory*) (LSTM) e redes neurais totalmente conectadas, juntamente com um pré-processamento substancial usando Equalização de histograma adaptativo limitada por contraste (do inglês *contrast-limited adaptive histogram equalization*) (CLAHE) para aprimoramento de imagem. [You et al., 2017] utilizaram as informações de localização de modo que o modelo fosse capaz de utilizar uma CNN para extração de recursos e uma RNN para aprendizado de modelo a fim de prever preços de imóveis. [Poursaeeda et al., 2018] avaliaram o impacto das características visuais de uma casa em seu valor de mercado utilizando redes neurais convolucionais profundas em um grande conjunto de dados de fotos de interiores e exteriores de residências para estimar o nível de luxo de fotos de imóveis, após isto desenvolveram também uma nova *framework* para avaliação automatizada de valor usando as fotos, além das características da casa, incluindo tamanho, preço oferecido e número de quartos. [Zhao et al., 2019] examinaram o uso de aprendizado profundo combinado com o *Extreme Gradient Boosting* (XGBoost) para avaliação imobiliária, analisando registros históricos de vendas junto com conteúdo visual, com fotos de casas online e pontuando cada imagem do ponto de vista estético para fazer uma previsão do preço da casa. [Wang et al., 2019] focaram nas imagens exteriores de imóveis para prever qual das duas imagens do mesmo imóvel é mais atrativa, para isto criaram um conjunto de dados e desenvolveram um modelo que garante valores de saída semelhantes para pares de imagens com atratividade semelhante e valores de saída diferentes para pares com atratividade diferente.

A principal contribuição deste trabalho é a análise de desempenho do algoritmo de CNN EfficientNetV2M [Tan e Le, 2021] em termos de acurácia, perda (MSE) e tempo de execução utilizando dois conjuntos de dados, o primeiro fornecido pelos autores de [Bappy et al., 2017] da Universidade da Califórnia e o segundo com incremento de imagens no primeiro conjunto de dados via buscas em ferramentas online para classificar as imagens em seis classes: banheiro, cozinha,

https://proceedings.science/p/157454?lang=pt-br

fachada, quarto, quintal e sala de estar. Por meio dessa análise de desempenho, há uma comparação entre o modelo EfficientNetV2M [Tan e Le, 2021] e os modelos EfficientNetB5 [Tan e Le, 2019] e ResNet50 [He et al., 2015].

Este artigo está organizado da seguinte forma: a seção 2 aborda a formulação do problema. A seção 3 visa discutir o conceito de EfficientNetV2M. A seção 4 apresenta os resultados das simulações computacionais. Por fim, a Seção 5 apresenta as principais conclusões.

2. Formulação do Problema

O mercado imobiliário brasileiro mantém-se aquecido desde 2019, depois de um período de crise que durou de 2014 a 2017. Este aquecimento está associado, principalmente, a alta disponibilidade de crédito e juros baixos. Segundo a Associação Brasileira de Incorporadoras Imobiliárias (Abrainc) e a Fipe [Abrainc, 2020], agosto de 2020 apresentou crescimento de unidades habitacionais comercializadas de $58,9\%$ para o mês na comparação com o mesmo período em 2019.

Outro fator que influenciou este crescimento foi a pandemia do COVID-19. Quando analisamos o cenário microeconômico, nota-se que a necessidade de realizar o isolamento social permitiu que as pessoas passassem mais tempo em suas residências, impulsionando venda, locação e reformas em imóveis para aumento do conforto e adaptação para um ambiente de *home-office*.

Diante deste cenário, o setor imobiliário começou a se digitalizar. Segundo a 4^a rodada da "Pesquisa da Influência do Coronavírus no Mercado Imobiliário Brasileiro" [DataZAP+, 2021] as maiores tendências a serem mantidas pós pandemia são a utilização de fotos profissionais, a disponibilidade de portfólio digitalizado e de fácil acesso em portais e/ou sites próprios, a realização de videochamadas e a assinatura de contratos digitais [Exame, 2021].

Conforme apontado por [You et al., 2017] as fotos podem facilmente dizer às pessoas exatamente qual a aparência do imóvel e podem ter uma sensação geral da casa, como por exemplo qual é o estilo geral de construção, como é o ambiente vizinho, etc.

3. EfficientNet V2M

Em 2021 a Google lançou o EfficientNetV2M [Tan e Le, 2021] que apresenta uma grande melhoria em relação ao EfficientNet [Tan e Le, 2019] em termos de velocidade de treinamento e melhor eficiência de parâmetros. Este modelo utiliza o conceito de aprendizado progressivo, o que significa que, embora os tamanhos das imagens sejam originalmente pequenos quando o treinamento começa, aumentam de tamanho progressivamente.

Na Tabela 1 encontra-se a arquitetura do modelo EfficientNetV2M. [Tan e Le, 2021] cita as distinções importantes comparado ao EfficientNet: (1) A primeira diferença é que o EfficientNetV2M usa amplamente MBConv ([Sandler et al., 2018]; [Tan e Le, 2019]) e o fused-MBConv ([Gupta e Tan, 2019]) nas camadas iniciais. (2) Em segundo lugar, EfficientNetV2 prefere taxas de expansão menores para MBConv, pois taxas de expansão menores tendem a ter menos sobrecarga de acesso à memória. (3) Em terceiro lugar, EfficientNetV2M prefere tamanhos de kernel $3x3$ menores, mas adiciona mais camadas para compensar o campo receptivo reduzido resultante do tamanho de kernel menor. (4) Por fim, o EfficientNetV2M remove completamente o último estágio stride-1 no EfficientNet original, devido ao seu grande tamanho de parâmetro e sobrecarga de acesso à memória.

O Fused-MBConv substitui o depthwise conv$3x3$ e o expansion conv$1x1$ no MBConv por um único conv$3x3$ regular nas camadas iniciais, conforme a Figura 1. Quando aplicado no estágio $1-3$ inicial, o Fused-MBConv pode melhorar a velocidade de treinamento com uma pequena sobrecarga nos parâmetros e FLOPs, mas se substituirmos todos os blocos por Fused-MBConv (estágio $1-7$), ele aumentará significativamente os parâmetros e FLOPs, além de diminuir a velocidade abaixo do treinamento [Tan e Le, 2021].

https://proceedings.science/p/157454?lang=pt-br

Tabela 1: Arquitetura do algoritmo EfficientNetV2M [Tan e Le, 2021].

Estágio	Operador	Passo	#Canais	#Camadas
0	Conv3x3	2	24	1
1	Fused-MBConv1, $k3x3$	1	24	2
2	Fused-MBConv4, $k3x3$	2	48	4
3	Fused-MBConv4, $k3x3$	2	64	4
4	MBConv4, $k3x3$, SE0.25	2	128	6
5	MBConv6, $k3x3$, SE0.25	1	160	9
6	MBConv6, $k3x3$, SE0.25	2	256	15
7	Conv1x1 & Pooling & FC	–	1280	1

Conforme descrito por [Tan e Le, 2021], no EfficientNet há o dimensionamento composto de todos os estágios (por exemplo, quando o coeficiente de profundidade é 2, todos os estágios nas redes duplicariam o número de camadas), porém esses estágios não contribuem igualmente para a velocidade de treinamento e eficiência dos parâmetros. No EfficientNetV2M, adota-se uma estratégia de dimensionamento não uniforme para adicionar gradualmente mais camadas aos estágios posteriores [Tan e Le, 2021]. Além disso, os EfficientNets aumentam agressivamente o tamanho da imagem, levando a um grande consumo de memória e treinamento lento, no EfficientNetV2M modifica-se ligeiramente a regra de dimensionamento e restringe-se o tamanho máximo da imagem a um valor menor [Tan e Le, 2021].

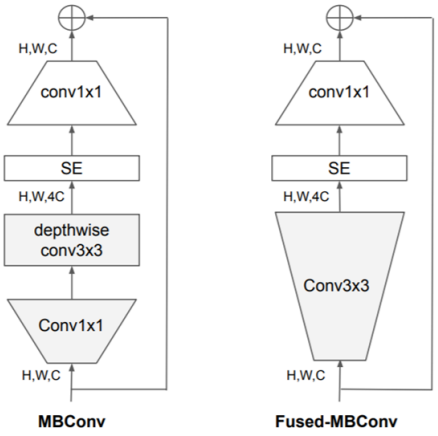


Figura 1: Estrutura do MBConv e do Fused-MBConv. [Tan e Le, 2021]

4. Resultados Experimentais

No presente trabalho foram utilizados 2 conjuntos de dados. O conjunto de dados denominado de original foi fornecido pelos autores do artigo [Bappy et al., 2017] da Universidade da Califórnia contendo 5859 imagens. O segundo conjunto de dados, denominado de atual, incrementa imagens no conjunto de dados original por meio de buscas em plataformas online, como Google, Pinterest, etc. O conjunto de dados atual contém 16391 imagens. Ambos apresentam 6 classes, sendo elas banheiro, cozinha, fachada, quarto, quintal e sala de estar.

Para a implementação dos conjuntos de dados na CNN, os conjuntos de dados foram divididos em conjuntos de treino e teste, sendo 80% das imagens para treino e 20% para teste. Os



(a) Banheiro (b) Cozinha (c) Fachada



(d) Quarto (e) Quintal (f) Sala de Estar

Figura 2: Exemplos de imagens das classes: (a) banheiro, (b) cozinha, (c) fachada, (d) quarto, (e) quintal e (f) sala de estar.

números de imagens de treino e teste dos conjuntos de dados em cada classe são mostrados na Tabela 2.

As métricas de avaliação do modelo utilizadas é a acurácia e o erro médio quadrático (do inglês *root squared error*) (MSE). A acurácia calcula a frequência com que as predições são iguais aos valores verdadeiros e é dada pela Equação 1. O MSE calcula a média da diferença entre o valor verdadeiro e o valor previsto ao quadrado.

$$\text{Acurácia} = \frac{\text{Número de predições corretas}}{\text{Número total de predições}} \tag{1}$$

Foram treinadas 3 arquiteturas de CNNs: a ResNet50, a EfficientNetB5 e a EfficientNetV2M. As arquiteturas foram pré-treinadas com o conjunto ImageNet e adicionadas camadas. Uma camada de convolução conv$3x3$ que cria um kernel(filtro) de convolução que se moverá por toda a imagem, da esquerda para a direita, de cima para baixo, aplicando um produto de convolução, onde a saída desta operação é chamada de imagem filtrada. Uma camada de normalização *BatchNormalization*. Uma camada de *Pooling*$2x2$ que reduz a amostra da entrada ao longo de suas dimensões espaciais (altura e largura) tomando o valor máximo em uma janela de entrada para cada canal da entrada. Uma camada para achatar a entrada (*flatten*). Uma camada *Dropout(*0.2*)* que descarta aleatoriamente ou omite unidades (ocultas e visíveis) durante o processo de treinamento visando a diminuição de *overfitting*. 3 camadas densas (*dense*) que é uma camada que está profundamente conectada com sua camada anterior, o que significa que os neurônios da camada estão conectados a todos os neurônios de sua camada anterior.

Todos os modelos foram treinados 10 vezes para cada conjunto de dados. A Tabela 3 exibe os resultados da performance de cada modelo em termos de média e desvio padrão.

É possível notar que a arquitetura EfficientNetV2M apresenta uma acurácia maior tanto para o conjunto de dados original quanto para o conjunto atual, com 93.42% e 90.59%, respectivamente. Ela também se destaca por ter menor MSE quando comparado as outras arquiteturas, 1.75%

https://proceedings.science/p/157454?lang=pt-br

Tabela 2: Número de imagens de treino e teste dos conjuntos de dados em cada classe.

Conjunto de Dados	Classe	Imagens para Treino	Imagens para Teste
Original	Banheiro	634	159
	Cozinha	794	198
	Fachada	707	177
	Quarto	1274	319
	Quintal	596	149
	Sala de Estar	682	170
Atual	Banheiro	2240	556
	Cozinha	2229	558
	Fachada	2264	554
	Quarto	2205	522
	Quintal	2000	500
	Sala de Estar	2215	548

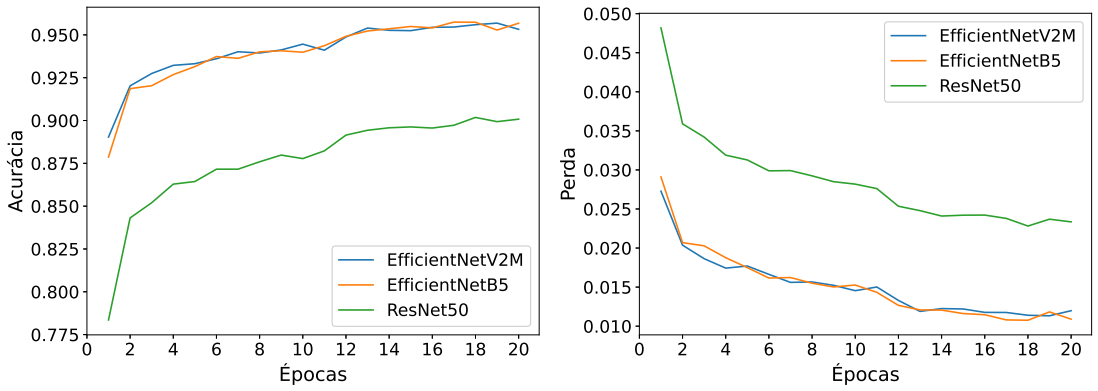
Tabela 3: Comparação de performance dos modelos em termos de média e desvio padrão.

Conjunto de Dados	Modelo	Acurácia (%)	Perda (%)	CPU (%)	Tempo decorrido (s)
Original	ResNet50	88.53 ± 2.34	3.47 ± 2.29	89.03 ± 1.17	4795 ± 160
	EfficientNetB5	92.83 ± 0.40	1.89 ± 0.10	83.88 ± 0.92	9325 ± 276
	EfficientNetV2M	93.42 ± 0.46	1.75 ± 0.12	87.20 ± 1.60	11158 ± 618
Atual	ResNet50	86.81 ± 1.01	3.64 ± 0.29	84.37 ± 15.24	7455 ± 550
	EfficientNetB5	89.36 ± 0.65	2.99 ± 0.20	79.34 ± 8.05	12166 ± 667
	EfficientNetV2M	90.59 ± 0.69	2.39 ± 0.13	84.42 ± 5.05	12375 ± 181

para o conjunto de dados original e 2.39% para o conjunto de testes atual. Percebe-se que o tempo de treinamento do ResNet50 é bem menor que os demais e nota-se também que o EfficientNetV2M, apesar de ter o maior tempo decorrido, altera-se menos que os demais quando comparado os conjuntos de dados, isto é, o número de imagens afeta menos no seu tempo decorrido que os demais modelos.

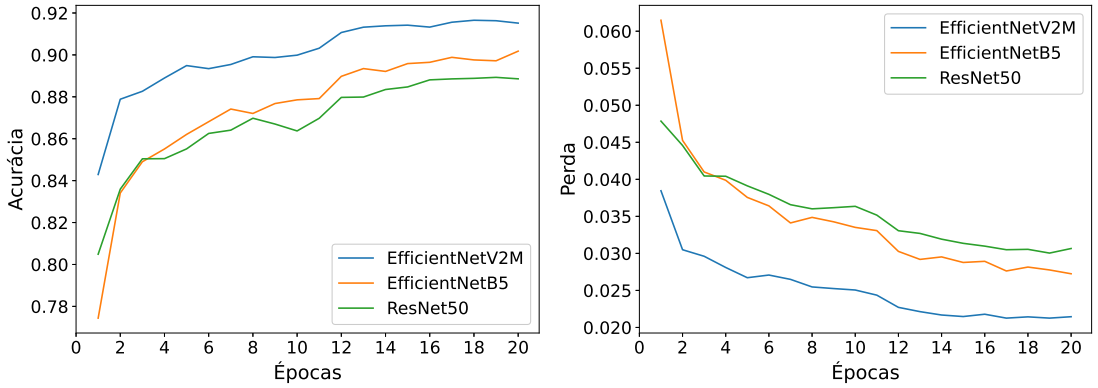
Na Figura 3 é mostrado um gráfico de época *versus* acurácia, descrevendo a média da acurácia dos treinamentos durante 20 épocas treinadas para cada modelo e um gráfico época *versus* perda, descrevendo a média do MSE dos treinamentos durante 20 épocas treinadas, ambos para o conjunto de dados original. Da mesma forma, na Figura 4 é mostrado o gráfico época *versus* acurácia e época *versus* perda do conjunto de dados atual.

Nota-se que nos gráficos de acurácia *versus* época, tanto para o conjunto de dados original (Figura 3a) quanto para o conjunto de dados atual (Figura 4a), que a acurácia tende a aumentar no decorrer das épocas para todos os 3 modelos, sendo que para o conjunto de dados original, os modelos EfficientNetB5 e EfficientNetV2M apresentam resultados semelhantes. O modelo ResNet50 por outro lado, apresenta uma curva com valores de acurácia piores que os outros modelos. Quando



(a) Gráfico acurácia *versus* época. (b) Gráfico perda *versus* época.

Figura 3: Gráficos (a) acurácia *versus* época e (b) perda *versus* época do conjunto de dados original.



(a) Gráfico acurácia *versus* época. (b) Gráfico perda *versus* época.

Figura 4: Gráficos (a) acurácia *versus* época e (b) perda *versus* época do conjunto de dados atual.

analisamos o gráfico acurácia *versus* época do conjunto de dados atual percebemos que as curvas são mais distintas umas das outras, sendo que o EfficientNetV2M possui a maior acurácia entre os modelos.

Analisando os gráficos de época *versus* perda, nota-se que o MSE tende a diminuir conforme o número de épocas. Na Figura 3b percebe-se que o ResNet50 possui maiores perdas, enquanto o EfficientNetB5 e o EfficientNetV2M possuem curvas com acurácias semelhantes. Já na Figura 4b, o modelo EfficientNetV2M apresenta perda bem menor que os demais modelos. Constata-se então, que o EfficientNetV2M tem um desempenho superior aos outros 2 modelos, o que exemplificado de forma mais clara quando se aumenta o conjunto de dados.

5. Conclusão

O presente trabalho analisou o desempenho do algoritmo EfficientNetV2M para a classificação de imagens imobiliárias. Para ambos os conjuntos de dados o EfficientNetV2M apresentou uma maior acurácia e uma menor perda. O tempo de execução do algoritmo é maior que os demais modelos mas apresenta uma menor diferença quando comparamos os resultados dos conjuntos de dados, o que nos leva a concluir que a quantidade de imagens interfere menos que nos demais modelos.

Trabalhos futuros serão focados no aumento do número de épocas do modelo para alcançar sua convergência.

Agradecimentos

Os autores gostariam de agradecer a empresa Vitrine Urbana Tecnologia Ltda. pelo suporte financeiro. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e da Universidade Federal de Juiz de Fora.

Referências

Abrainc (2020). Vendas de imóveis crescem 58,9% em agosto, após novo recorde. https://www.abrainc.org.br/abrainc/releases/2020/10/29/vendas-de-imoveis-crescem-589-em-agosto-apos-novo-recorde/. Acessado: 2022-05-10.

Albawi, S., Mohammed, T. A., e Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, p. 1–6.

Arel, I., Rose, D., e Karnowski, T. (2010). Deep machine learning - a new frontier in artificial intelligence research [research frontier]. *IEEE Comp. Int. Mag.*, 5:13–18.

Bappy, M. J., Barr, J., Srinivasan, N., e Roy-Chowdhury, A. (2017). Real estate image classification. *IEEE Winter Conference on Applications of Computer Vision (WACV)*.

DataZAP+ (2021). A influência do coronavírus no mercado imobiliário brasileiro. https://imagens.vivareal.com.br/zapmais/movimento-vaidecorretor/datazap-abr2021-4-onda.pdf. Acessado: 2022-05-10.

Exame (2021). O futuro do mercado imobiliário em um mundo pós-pandemia. https://exame.com/colunistas/genoma-imobiliario/o-futuro-do-mercado-imobiliario-em-um-mundo-pos-pandemia/. Acessado: 2022-05-10.

Gupta, S. e Tan, M. (2019). Efficientnet-edgetpu: Creating accelerator-optimized neural networks with automl. In *Google AI Blog*. URL https://ai.googleblog.com/2019/08/efficientnet-edgetpu-creating.html.

Géron, A. (2021). *Mãos à Obra: Aprendizado de Máquina com Scikit-Learn. Keras & TensorFlow*. Alta Books Editora, Rio de Janeiro. ISBN 978-85-5081-548-0.

He, K., Zhang, X., Ren, S., e Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, abs/1512.03385. URL http://arxiv.org/abs/1512.03385.

Indolia, S., Goswami, A. K., Mishra, S., e Asopa, P. (2018). Conceptual understanding of convolutional neural network- a deep learning approach. *Procedia Computer Science*, 132:679–688. ISSN 1877-0509. URL https://doi.org/10.1016/j.procs.2018.05.069. International Conference on Computational Intelligence and Data Science.

LeCun, Y., Bengio, Y., e Hinton, G. (2015). Deep learning. *Nature*, 521:436–444. URL http://dx.doi.org/10.1038/nature14539.

https://proceedings.science/p/157454?lang=pt-br

LeCun, Y., Bottou, L., Bengio, Y., e Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the Institute of Radio Engineers*, 86(11):2278–2323. ISSN 0018-9219.

Poursaeeda, O., Matera, T., e Belongie, S. (2018). Vision-based real estate price estimation. *Machine Vision and Applications*, 29.

Rosa, F. D., Favretto, L. D., Rodrigues, V. B., e Gharaibeh, N. G. (2020). Avaliação do desempenho de redes neurais convolucionais para detecção automática de trincas em pavimentos. *Transportes (Rio de Janeiro)*, 28(5). ISSN 2237-1346.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., e Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Computer Vision and Pattern Recognition*.

Smirnov, E., Timoshenko, D., e Andrianov, S. (2014). Comparison of regularization methods for imagenet classification with deep convolutional neural networks. *AASRI Procedia*, 6:89–94.

Tan, M. e Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946. URL `http://arxiv.org/abs/1905.11946`.

Tan, M. e Le, Q. V. (2021). Efficientnetv2: Smaller models and faster training. *CoRR*, abs/2104.00298. URL `https://arxiv.org/abs/2104.00298`.

Tivive, F. H. C. e Bouzerdoum, A. (2005). Efficient training algorithms for a class of shunting inhibitory convolutional neural networks. *Neural Networks, IEEE Transactions on*, 16:541 – 556.

Wang, X., Takada, Y., Kado, Y., e Yamasaki, T. (2019). Predicting the attractiveness of real-estate images by pairwise comparison using deep learning. In *2019 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, p. 84–89.

You, Q., Pang, R., Cao, L., e Luo, J. (2017). Image-based appraisal of real estate properties. *IEEE Transactions on Multimedia*, 19(12):2751–2759.

Zhao, Y., Chetty, G., e Tran, D. (2019). Deep learning with xgboost for real estate appraisal. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, p. 1396–1401.

https://proceedings.science/p/157454?lang=pt-br