Classificação automática de imagens utilizando redes neurais convolucionais MobileNet: um estudo de caso com reconhecimento de felinos

Vitor Amadeu Souza[1]; 0009-0002-1857-6799

1 – UniFOA, Centro Universitário de Volta Redonda, Volta Redonda, RJ.
vitor.amadeu@foa.org.br

Resumo: Este trabalho apresenta uma implementação da arquitetura MobileNet para classificação automática de imagens, com foco específico no reconhecimento de felinos domésticos. O estudo utilizou um modelo MobileNet pré-treinado na base de dados ImageNet, aplicando-o na classificação de uma imagem de um gato persa. A metodologia envolveu o carregamento e pré-processamento da imagem, seguido pela aplicação do modelo de deep learning para obtenção das predições. Os resultados demonstraram precisão na classificação, com 90,97% de confiança para a classe "Persian_cat", evidenciando a eficácia das redes neurais convolucionais móveis para tarefas de visão computacional. O estudo confirma a aplicabilidade das arquiteturas MobileNet em cenários reais de classificação de imagens, oferecendo uma solução computacionalmente eficiente para dispositivos com recursos limitados. A pesquisa contribui para a compreensão das capacidades e limitações dos modelos de deep learning em tarefas de reconhecimento visual automatizado.

Palavras-chave: MobileNet; Redes Neurais Convolucionais. Classificação de Imagens. Deep Learning. Visão Computacional. ImageNet.

INTRODUÇÃO

A classificação automática de imagens representa um dos principais desafios na área de visão computacional, tendo experimentado avanços significativos com o desenvolvimento das redes neurais convolucionais (CNNs) profundas. Segundo Krizhevsky *et al.* (2012), a introdução da arquitetura AlexNet marcou um ponto de inflexão no campo, demonstrando a superioridade das CNNs em tarefas de reconhecimento visual. Desde então, diversas arquiteturas foram propostas, buscando melhorar tanto a precisão quanto a eficiência computacional dos modelos.

A arquitetura MobileNet, proposta por Howard *et al*. (2017), surge como uma solução inovadora para o dilema entre precisão e eficiência computacional em modelos de deep learning. Os autores introduziram o conceito de convoluções separáveis em profundidade (depthwise separable convolutions), que reduz significativamente o número de parâmetros e operações computacionais sem comprometer substancialmente a performance. Esta abordagem tornou-se particularmente relevante para aplicações em dispositivos móveis e sistemas embarcados, onde os recursos computacionais são limitados.

O processo de classificação de imagens utilizando CNNs envolve múltiplas etapas, desde o pré-processamento dos dados até a interpretação dos resultados finais. Deng *et al*. (2009) destacam a importância da base de dados ImageNet como benchmark padrão para avaliação de modelos de classificação de imagens, contendo mais de 14 milhões de imagens distribuídas em milhares de categorias. A utilização de modelos pré-treinados nesta base de dados permite a aplicação de transfer learning, técnica que possibilita a reutilização de conhecimento previamente adquirido em novas tarefas.

A classificação específica de felinos domésticos apresenta desafios únicos devido à variabilidade intraclasse e às semelhanças interclasses existentes entre diferentes raças. Zhang (2014) demonstrou que as características morfológicas dos felinos, como padrões de pelagem, formato facial e estrutura corporal, podem ser efetivamente capturadas por CNNs modernas. O reconhecimento automático de raças felinas tem aplicações práticas em diversas áreas, incluindo medicina veterinária, sistemas de identificação de animais perdidos e aplicações comerciais relacionadas ao mercado pet.

A evolução das técnicas de pré-processamento de imagens também desempenha papel vital na performance dos modelos de classificação. Shorten e Khoshgoftaar (2019) enfatizam a importância da normalização e padronização dos dados de entrada, processos que garantem a convergência adequada dos algoritmos de otimização durante o treinamento. No contexto das MobileNets, o pré-processamento específico proposto pelos autores originais inclui normalização de pixels para o intervalo [-1, 1], otimizando a representação das características visuais.

O presente estudo busca avaliar a eficácia da arquitetura MobileNet na classificação de imagens de felinos, contribuindo para o corpo de conhecimento sobre a aplicabilidade de modelos de deep learning em cenários reais. A pesquisa também visa demonstrar a facilidade de implementação de soluções baseadas em MobileNet, fornecendo insights sobre as capacidades e limitações desta arquitetura em tarefas específicas de reconhecimento visual.

MÉTODOS

A metodologia adotada neste estudo seguiu uma abordagem experimental quantitativa, utilizando um modelo MobileNet pré-treinado para classificação de uma imagem específica de felino doméstico. A implementação foi realizada utilizando a linguagem Python, com as bibliotecas TensorFlow 2.0 e Keras para manipulação dos modelos de deep learning, conforme recomendado por Abadi *et al*. (2016) para implementações eficientes de redes neurais.

O dataset de referência utilizado foi o ImageNet, conforme estabelecido por Russakovsky *et al*. (2015), que constitui o padrão para avaliação de modelos de classificação de imagens em larga escala. O modelo MobileNet empregado foi pré-treinado nesta base de dados, permitindo a aplicação direta de transfer learning sem necessidade de retreinamento específico para a tarefa proposta.

A imagem utilizada como objeto de estudo foi obtida através de uma requisição HTTP à plataforma Wikimedia Commons, especificamente uma fotografia de um gato da raça persa. A escolha desta imagem foi motivada pela disponibilidade pública e pela qualidade visual adequada para teste de modelos de classificação. O processo de aquisição da imagem

incluiu a configuração de cabeçalhos HTTP apropriados, incluindo User-Agent, para garantir o acesso adequado ao recurso web, seguindo as melhores práticas descritas por Fielding *et al.* (1999) para requisições HTTP.

O pré-processamento da imagem seguiu as especificações da arquitetura MobileNet, conforme definido por Howard *et al*. (2017). Inicialmente, a imagem foi redimensionada para as dimensões 224x224 pixels, requisito padrão para entrada na rede neural. Posteriormente, os valores dos pixels foram convertidos para array NumPy e normalizados utilizando a função preprocess_input específica da MobileNet, que aplica normalização no intervalo [-1, 1]. Este processo de normalização é essencial para a performance adequada do modelo, como demonstrado por Ioffe e Szegedy (2015) em seus estudos sobre normalização de batches.

A inferência foi realizada através do método predict do modelo, gerando probabilidades para cada uma das 1000 classes do ImageNet. Os resultados foram decodificados utilizando a função decode_predictions, que mapeia os índices das classes para suas respectivas labels textuais e organiza os resultados em ordem decrescente de probabilidade.

A análise dos resultados focou nas três principais predições geradas pelo modelo, avaliando tanto a precisão da classificação quanto a interpretabilidade dos resultados obtidos. Esta abordagem top-k é amplamente utilizada na literatura de classificação de imagens, conforme demonstrado por He *et al*. (2016) em seus estudos sobre ResNet.

O código-fonte desta pesquisa está disponível para download através do link: https://github.com/vitor-souza-ime/mobilenet.

RESULTADOS E DISCUSSÃO

A aplicação do modelo MobileNet na imagem de teste produziu bons resultados em termos de precisão e especificidade da classificação. A análise das três principais predições revelou que o modelo identificou corretamente a classe "Persian_cat" com probabilidade de 90,97%, demonstrando alta confiança na classificação. Esta performance alinha-se com os resultados reportados por Howard *et al*. (2017) em seus experimentos originais com a

arquitetura MobileNet, onde os autores documentaram precisões superiores a 89% no dataset ImageNet.

A segunda predição mais provável foi "Pomeranian" com 5,81% de probabilidade, resultado que, embora incorreto, revela aspectos interessantes sobre o funcionamento interno da rede neural. A confusão entre um felino e um canino de pequeno porte sugere que o modelo está capturando características visuais relacionadas à textura da pelagem e tamanho relativo do animal na imagem. Este fenômeno é consistente com os achados de Zeiler e Fergus (2014), que demonstraram através de técnicas de visualização que CNNs em camadas intermediárias frequentemente respondem a texturas e padrões locais antes de formar representações de alto nível.

A terceira predição, "park_bench" com 0,64% de probabilidade, indica que o modelo também está detectando elementos do background da imagem, especificamente o banco de madeira onde o gato está posicionado. Esta capacidade de reconhecimento contextual é uma característica emergente das CNNs profundas, conforme documentado por Girshick *et al.* (2014) em seus estudos sobre detecção de objetos. A baixa probabilidade associada a esta classe demonstra que o modelo está adequadamente focando no objeto principal da imagem. A Figura 1 apresenta a imagem usada para fazer a predição.

A distribuição das probabilidades obtidas segue um padrão típico de modelos bem treinados, com uma classe dominante e probabilidades rapidamente decrescentes para as demais classes. Este comportamento indica que a função softmax está operando adequadamente, produzindo uma distribuição de probabilidades bem calibrada, aspecto fundamental para aplicações práticas conforme destacado por Guo *et al.* (2017) em seus estudos sobre calibração de redes neurais.

A alta precisão obtida na classificação do gato persa pode ser atribuída a diversos fatores metodológicos e arquiteturais. Primeiramente, a qualidade da imagem utilizada, com iluminação adequada e foco nítido no objeto de interesse, contribui para a performance do modelo. Dodge e Karam (2016) demonstraram como a qualidade da imagem impacta diretamente na precisão de modelos de classificação, especialmente em cenários com ruído ou distorções.

Figura 1 - Imagem usada para predição

Fonte: WATKIN, A (Wikipedia)

A arquitetura MobileNet demonstrou sua eficiência computacional característica durante a execução do experimento. O tempo de inferência observado foi inferior a 500ms em hardware convencional, confirmando a adequação desta arquitetura para aplicações em tempo real. Esta eficiência deriva das convoluções separáveis em profundidade, que reduzem o número de operações matemáticas necessárias sem comprometer a capacidade representacional do modelo, conforme quantificado por Sandler *et al*. (2018) na versão MobileNetV2.

CONCLUSÕES

O presente estudo demonstrou a eficácia da arquitetura MobileNet para classificação automática de imagens de felinos domésticos, alcançando precisão de 90,97% na identificação correta da classe "Persian_cat". Os resultados obtidos confirmam as capacidades reportadas na literatura original da arquitetura, validando sua aplicabilidade em

cenários reais de reconhecimento visual. A implementação realizada evidenciou o uso de modelos pré-treinados utilizando as APIs disponíveis no ecossistema TensorFlow/Keras, contribuindo para o acesso a tecnologias de deep learning.

Os resultados obtidos sugerem direções promissoras para pesquisas futuras, incluindo a avaliação da performance em datasets mais amplos e diversificados, a investigação de técnicas de fine-tuning para domínios específicos e a exploração de arquiteturas mais recentes da família MobileNet. A metodologia apresentada pode ser adaptada para outras tarefas de classificação, contribuindo para a expansão das aplicações práticas de deep learning em diversos domínios.

REFERÊNCIAS

ABADI, Martín et al. TensorFlow: A system for large-scale machine learning. In: 12th USENIX symposium on operating systems design and implementation (OSDI 16). 2016. p. 265-283. Disponível em: https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf. Acesso em: 14 set. 2025.

DENG, Jia et al. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. IEEE, 2009. p. 248-255. DOI: 10.1109/CVPR.2009.5206848. Acesso em: 14 set. 2025.

DODGE, Samuel; KARAM, Lina. Understanding how image quality affects deep neural networks. In: 2016 eighth international conference on quality of multimedia experience (QoMEX). IEEE, 2016. p. 1-6. DOI: 10.1109/QoMEX.2016.7498955. Acesso em: 14 set. 2025.

FIELDING, Roy et al. Hypertext transfer protocol--HTTP/1.1. RFC 2616, 1999. Disponível em: https://tools.ietf.org/html/rfc2616. Acesso em: 14 set. 2025.

GIRSHICK, Ross et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. p. 580-587. DOI: 10.1109/CVPR.2014.81. Acesso em: 14 set. 2025.

GUO, Chuan et al. On calibration of modern neural networks. In: International conference on machine learning. PMLR, 2017. p. 1321-1330. Disponível em: http://proceedings.mlr.press/v70/guo17a.html. Acesso em: 14 set. 2025.

HE, Kaiming et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778. DOI: 10.1109/CVPR.2016.90. Acesso em: 14 set. 2025.

HOWARD, Andrew G. et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861, 2017. DOI: 10.48550/arXiv.1704.04861. Acesso em: 14 set. 2025.

IOFFE, Sergey; SZEGEDY, Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. PMLR, 2015. p. 448-456.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. Communications of the ACM, v. 60, n. 6, p. 84-90, 2017. DOI: 10.1145/3065386. Acesso em: 14 set. 2025.

RUSSAKOVSKY, Olga et al. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, v. 115, n. 3, p. 211-252, 2015. DOI: 10.1007/s11263-015-0816-y. Acesso em: 14 set. 2025.

SANDLER, Mark et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 4510-4520. DOI: 10.1109/CVPR.2018.00474. Acesso em: 14 set. 2025.

SHORTEN, Connor; KHOSHGOFTAAR, Taghi M. A survey on image data augmentation for deep learning. Journal of big data, v. 6, n. 1, p. 1-48, 2019.

ZEILER, Matthew D.; FERGUS, Rob. Visualizing and understanding convolutional networks. In: European conference on computer vision. Springer, 2014. p. 818-833. DOI: 10.1007/978-3-319-10590-1_53. Acesso em: 14 set. 2025.

ZHANG, Ning et al. Part-based R-CNNs for fine-grained category detection. In: European conference on computer vision. Springer, 2014. p. 834-849. DOI: 10.1007/978-3-319-10590-1_54. Acesso em: 14 set. 2025.

WATKIN, A. Gato. Wikipédia, a enciclopédia livre, 24 ago. 2025. Disponível em: https://pt.wikipedia.org/wiki/Gato. Acesso em: 15 set. 2025.