

Revolutionizing Security: An Analysis of Convolutional Neural Networks in Image Detection

Pedro A. Alves Viana Cotta

¹Departamento de Computação – Universidade Federal de Ouro Preto (UFOP)
Ouro Preto – MG – Brazil

pedro.cotta@aluno.ufop.edu.br

Abstract. *This article presents the use of convolutional neural networks (CNNs) in the area of security and surveillance, focusing on image detection in security cameras. The importance of video surveillance, the challenges of manual analysis of security videos and the need for automation through CNNs are discussed. A CNN architecture is applied to the CIFAR-10 dataset for classifying objects in color images, demonstrating the effectiveness of the approach. Related work highlights the use of CNNs in various security and surveillance applications, while notable architectures such as VGGNet and ResNet exemplify the significant advances made in the area. These results highlight the potential of CNNs to improve the efficiency and accuracy of security and surveillance systems, contributing to smarter and more effective monitoring.*

Resumo. *Este artigo apresenta o uso de redes neurais convolucionais (CNNs) na área de segurança e vigilância, com foco na detecção de imagens em câmeras de segurança. Aborda-se a importância da vigilância por vídeo, os desafios da análise manual de vídeos de segurança e a necessidade de automação por meio de CNNs. Uma arquitetura de CNN é aplicada ao conjunto de dados CIFAR-10 para classificação de objetos em imagens coloridas, demonstrando a eficácia da abordagem. Trabalhos relacionados destacam o uso das CNNs em diversas aplicações de segurança e vigilância, enquanto arquiteturas notáveis como VGGNet e ResNet exemplificam os avanços significativos alcançados na área. Esses resultados destacam o potencial das CNNs em aprimorar a eficiência e a precisão dos sistemas de segurança e vigilância, contribuindo para um monitoramento mais inteligente e eficaz.*

1. Introdução

A vigilância por vídeo desempenha um papel crucial na segurança pública e privada, permitindo a monitorização de locais em tempo real e a análise retrospectiva de eventos. Em ambientes como áreas urbanas, instalações industriais e de transporte, a vigilância por vídeo é essencial para a detecção e prevenção de crimes, monitoramento de multidões e segurança de infraestruturas críticas. No entanto, a análise manual de vídeos de segurança é trabalhosa e suscetível a erros humanos, especialmente quando se trata de grandes volumes de dados. A identificação de eventos relevantes, como atividades suspeitas ou incidentes, pode ser difícil e demorada, levando a atrasos na resposta e comprometendo a eficácia da vigilância.

As redes neurais têm desempenhado um papel cada vez mais importante na detecção de imagem em vídeos de segurança. Essas redes são capazes de aprender padrões complexos em dados visuais e podem ser treinadas para reconhecer objetos, pessoas e comportamentos específicos em vídeos, automatizando a análise e identificação de eventos relevantes.

Este artigo aborda o tema da segurança e vigilância com detecção de imagens por meio de redes neurais convolucionais (CNNs). O objetivo é criar um modelo ou algoritmo com a capacidade de classificar diversas imagens, com foco na aplicação prática em câmeras de segurança e vigilância. Para ilustrar essa capacidade, aplicamos uma arquitetura de CNN a um conjunto de dados do CIFAR-10, que consiste em 60.000 imagens coloridas de 32x32 pixels, divididas em 10 classes diferentes. Essa aplicação permite avaliar a eficácia da CNN em reconhecer e classificar objetos comuns em diversos cenários, fornecendo insights valiosos para aprimorar sistemas de segurança e vigilância baseados em imagens.

Além disso, exploramos trabalhos relacionados que destacam o uso das CNNs em segurança e vigilância. Pesquisas como a proposta de um Modelo de Rede Neural Convolucional para sistemas de vigilância remota móvel, o desenvolvimento de um Sistema Aprimorado de Controle Residencial Inteligente e Segurança baseado em aprendizagem profunda, a introdução das Redes Neurais Convolucionais Restritas para detecção de manipulação de imagens, e o reconhecimento facial baseado em CNNs são exemplos significativos do impacto dessas tecnologias nessas áreas. Esses trabalhos mostram o potencial das CNNs em melhorar a eficiência e precisão dos sistemas de segurança e vigilância, tornando-os mais adaptáveis e eficazes em diversas aplicações.

Por fim, discutimos arquiteturas notáveis de CNNs, como a VGGNet, a ResNet e o DeepFace, que mostraram avanços significativos em suas áreas de aplicação, ampliando o uso das CNNs em diferentes contextos de segurança e vigilância. Esses avanços destacam o potencial das redes neurais convolucionais em melhorar a segurança e a vigilância através da análise de imagens, contribuindo para sistemas mais inteligentes e eficazes.

2. Trabalhos relacionados

O uso de Redes Neurais Convolucionais (CNNs) tem sido objeto de intensa pesquisa na área de segurança e vigilância, destacando sua eficácia em diversas tarefas. [Ayad and Al-Khalidi 2023] propuseram um Modelo de Rede Neural Convolucional (CNN) para sistema de vigilância remota móvel, evidenciando a capacidade das CNNs em lidar com imagens de forma eficaz para segurança doméstica.

[Taiwo et al. 2022] apresentaram um Sistema Aprimorado de Controle Residencial Inteligente e Segurança baseado em aprendizagem profunda, demonstrando a aplicação prática das CNNs no controle residencial e segurança, ampliando o uso desses modelos em ambientes domésticos.

[Bayar and Stamm 2018] introduziram as Redes Neurais Convolucionais Restritas (R-CNNs), uma abordagem inovadora para detecção de manipulação de imagens, evidenciando como as CNNs podem ser adaptadas para resolver problemas específicos de segurança da informação, ampliando a aplicabilidade desses modelos em diferentes contextos.

[Coşkun et al. 2017] exploraram o Reconhecimento Facial baseado em Rede Neural Convolutacional, destacando a capacidade das CNNs em reconhecer faces com precisão, mesmo em condições desafiadoras, ampliando as possibilidades de aplicação desses modelos em sistemas de segurança.

Trabalhos pioneiros como o de [Krizhevsky et al. 2012] demonstraram a eficácia das CNNs na Classificação ImageNet, estabelecendo as bases para o uso generalizado desses modelos em tarefas de visão computacional, ampliando o entendimento sobre o potencial dessas redes.

Arquiteturas notáveis como a VGGNet de [Simonyan 2014], a ResNet de [He et al. 2016] e o DeepFace de [Taigman et al. 2014] mostraram avanços significativos em suas áreas de aplicação, ampliando o uso das CNNs em diferentes contextos de segurança e vigilância.

O trabalho de [Hasan et al. 2016] destacou a importância de aprender regularidades temporais em sequências de vídeo, essencial para sistemas de vigilância baseados em CNNs, ampliando a compreensão sobre a aplicabilidade desses modelos em análise de comportamento.

Esses avanços destacam o potencial das Redes Neurais Convolucionais (CNNs) em melhorar a segurança e a vigilância através da análise de imagens. A pesquisa de [Bayar and Stamm 2018] sobre Redes Neurais Convolucionais Restritas (R-CNNs) é especialmente relevante, introduzindo uma abordagem inovadora para a detecção de manipulação de imagens, contribuindo para a segurança da informação em ambientes digitais.

O reconhecimento facial, uma área essencial em segurança, tem sido significativamente aprimorado com o uso de CNNs. O trabalho de [Coşkun et al. 2017] destaca os avanços nesse campo, com aplicações que vão desde a identificação de pessoas em imagens até a autenticação biométrica.

Além disso, o reconhecimento de padrões temporais em sequências de vídeo, como destacado por [Hasan et al. 2016], é fundamental para identificar comportamentos suspeitos em sistemas de vigilância. Essa capacidade é essencial para garantir a segurança em ambientes dinâmicos e complexos.

Com isso, pode-se dizer que, o uso de Redes Neurais Convolucionais tem revolucionado a segurança e a vigilância, proporcionando avanços significativos em áreas como detecção de objetos, reconhecimento facial e análise de comportamento. Esses avanços demonstram o potencial das CNNs em melhorar a eficiência e a precisão dos sistemas de segurança, tornando-os mais adaptáveis e eficazes em diversas aplicações.

3. Metodologia

As redes neurais convolucionais (CNNs) são amplamente utilizadas na detecção de imagens. Para ilustrar a capacidade das câmeras de segurança e vigilância de identificar diferentes tipos de imagens em vários cenários, aplicaremos uma arquitetura de CNN a um conjunto de dados do CIFAR-10. Este conjunto de dados consiste em 60.000 imagens coloridas de 32x32 pixels, divididas em 10 classes diferentes, como aviões, carros, pássaros, gatos, entre outros. O uso do CIFAR-10 permite avaliar a capacidade da CNN

em reconhecer e classificar objetos comuns em ambientes diversos, fornecendo insights valiosos sobre a eficácia da abordagem em aplicações práticas de segurança e vigilância.

Antes de alimentar as imagens à CNN, é realizado um pré-processamento para normalização e redimensionamento. Normaliza as intensidades dos pixels para o intervalo $[0,1]$ dividindo todos os valores de pixel por 255. Com isso, é possível ter uma melhor estabilidade no treinamento da rede. Além disso, é redimensionado todas as imagens para o tamanho esperado pela rede, neste caso, 32x32 pixels.

A arquitetura da CNN utilizada neste estudo é composta por várias camadas. A primeira camada é uma camada de convolução, seguida por camadas de ativação ReLU e camadas de pooling. Este padrão se repete várias vezes, com algumas variações na arquitetura para melhor se adequar aos dados do CIFAR-10. A arquitetura final é a seguinte:

1. Primeira camada convolucional: Camada de Convolução, com 64 filtros de tamanho 3x3, entrada de forma ativação ReLU para ter uma eficácia e eficiência computacional, preenchimento '*same*', entrada de forma (32,32,3) para as imagens CIFAR 10 (32x32 pixels com 3 canais de cor para vermelho, verde e azul), uma camada de Pooling para uma operação de subamostragem que reduz a dimensionalidade das representações, mantendo as características mais importantes, MaxPooling com tamanho 2x2.
2. Segunda camada convolucional: Camada de Convolução, com 128 filtros de tamanho 3x3, entrada de forma ativação ReLU, preenchimento '*same*', uma camada de Pooling, MaxPooling com tamanho 2x2.
3. Terceira camada convolucional: Camada de Convolução, com 256 filtros de tamanho 3x3, entrada de forma ativação ReLU, preenchimento '*same*', uma camada de Pooling, MaxPooling com tamanho 2x2.
4. Camada Densa com 10 neurônios, ativação ReLU, Cada neurônio nesta camada representa uma classe no conjunto de dados CIFAR-10, pois o conjunto de dados tem 10 classes diferentes
5. Camada de Dropout: Probabilidade de dropout de 0.25.
6. Camada de Saída: 10 neurônios (um para cada classe), ativação softmax.

O modelo foi compilado com a função de perda '*sparse_categorical_crossentropy*' e o otimizador '*adam*'. Foi treinado por 10 épocas com um tamanho de lote de 32 imagens e uma divisão de validação de 0.1 do conjunto de treinamento.

Após o treinamento, o modelo foi avaliado no conjunto de teste para determinar sua precisão na classificação das imagens.

4. Resultados

Os resultados do treinamento da rede neural convolucional (CNN) no conjunto de dados CIFAR-10 mostraram uma melhoria progressiva na precisão e na redução da perda ao longo das épocas. Inicialmente, a precisão era baixa, mas aumentou significativamente nas primeiras épocas, alcançando cerca de 88% de precisão no conjunto de treinamento e 75% no conjunto de validação após 10 épocas. A perda diminuiu consistentemente durante o treinamento, indicando que o modelo estava aprendendo com sucesso os padrões nos dados. No entanto, houve sinais de leve sobreajuste nas últimas épocas, visto que a precisão no conjunto de validação começou a estagnar e a perda a aumentar ligeiramente.

Na avaliação final, o modelo obteve uma precisão de cerca de 75% no conjunto de teste, demonstrando sua capacidade de generalização para novos dados. Esses resultados sugerem que a CNN é eficaz na classificação de imagens do CIFAR-10, embora possa se beneficiar de técnicas adicionais de regularização para melhorar o desempenho em dados não vistos.

References

- Ayad, R. and Al-Khalidi, F. Q. (2023). Convolutional neural network (cnn) model to mobile remote surveillance system for home security. *International Journal of Computing and Digital Systems*, 13(1):1–1.
- Bayar, B. and Stamm, M. C. (2018). Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection. *IEEE Transactions on Information Forensics and Security*, 13(11):2691–2706.
- Coşkun, M., Uçar, A., Yildirim, Ö., and Demir, Y. (2017). Face recognition based on convolutional neural network. In *2017 international conference on modern electrical and energy systems (MEES)*, pages 376–379. IEEE.
- Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A. K., and Davis, L. S. (2016). Learning temporal regularity in video sequences. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 733–742.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Identity mappings in deep residual networks. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 630–645. Springer.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Simonyan, Karen e Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.
- Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708.
- Taiwo, O., Ezugwu, A. E., Oyelade, O. N., and Almutairi, M. S. (2022). Enhanced intelligent smart home control and security system based on deep learning model. *Wireless communications and mobile computing*, 2022:1–22.