Computer Vision Artificial Intel- ITAI 1378

Khanh Huynh

Prof. Patricia McManus

10/04/2024

|  |  |
| --- | --- |
| CNN |  |

1. **CNN Architectures**

In this context, CNNs and traditional neural networks are a case of two very divergent methodologies embraced in deep learning, especially in the setting of image data. CNNs are designed solely for grid-like data structures such as images, while standard neural networks, or fully connected networks, can have a variety of general uses, assuming uniform treatment for each input variable, regardless of spatial correlations.

Evidently, these two architectures have varying strengths and weaknesses when it comes to visual tasks relating to image classification. A CNN has several different key layers, all working to represent distinct characteristics of visual information. The convolutional layer is one of the crucial units of CNNs; they are small trainable kernels, the sizes of which are typically 3x3 or 5x5. They sweep across the input images and try to detect a neighboring ensemble of characteristics like edges, corners, and textures. As the layer progresses deep into the network, high-level, more abstract features are created from low-level visual features, thus forming a gradual hierarchy of complex visual patterns. Normally, the pooling layers come after convolutional layers, where the pooling operations reduce the scale of the feature maps using max pooling methods to reduce computation complexity while maintaining the important features. As a whole, CNNs contain fully connected layers even at the end, where high-level features are fused to generate the final output. The CNNs apply an activation function like ReLU after every convolution operation to introduce non-linearity, and thus, allow for high-level representations to come together seamlessly in terms of functioning.

In contrast, a classical model consists solely of fully connected layers, wherein neurons in a layer attach completely to neurons in the next layer. This network in its structure afterwards flattens the input so that ultimately an image is being expressed as a one-dimensional vector, which in turn sacrifices any notions of spatial correlations among pixels. This flattening process makes traditional networks treat every pixel as an independent input. The spatial redundancy in image data, such as local patterns like edges and textures, is not explicitly taken into account when images are input into these networks. The inability to capture such spatial structure in the input is a serious limitation when the network is used on tasks such as image classification. More specifically, traditional networks are not able to capture local patterns and general spatial hierarchies present inside the images. To get around this limitation, they need a much higher number of parameters to explicitly capture spatial dependencies in the input. The major difference between the two kinds of networks can be understood with regard to spatial data management. Conventional networks are not so good at capturing spatial hierarchies and local patterns present inside the images, which affects generalization performance. This makes them suitable for working with higher dimensional inputs like images.

On the other hand, Convolutional Neural Networks, or CNNs, are good at efficiently managing spatial hierarchies and local patterns present inside the images. This is primarily due to the convolutional and pooling layers employed in them. As a direct result, CNNs significantly outperform conventional networks in an image-related task, as was demonstrated by the "Chihuahua versus muffin" classifier described above. This is because CNNs are very good at learning from local patterns inside images, while traditional methods are not. Moreover, CNNs can learn to do so in a much more efficient manner than traditional methods. They can decrease dimensionality and concentrate on learning features, without having to use an enormous number of parameters.

To sum up, while conventional neural networks are supposed to be quite good in performance on a wide array of tasks, they are not designed to be good in learning from an image. On the other hand, Convolutional Neural Networks are designed to learn from an image by maintaining spatial relations thanks to their convolutional and pooling layers. As explained earlier, CNNs are the optimal choice for working with visual data due to their ability to more efficiently learn from local patterns in images compared to other methodologies.

1. **Model Performance**

The CNN model we constructed is able to attain a success rate of 85% on the image classification task "Chihuahua vs Muffin" from the test set. Therefore, if one were to show it, say, 100 pictures, it would make a correct guess 85 times, either saying it is a chihuahua or that it is a muffin on average. The remaining 15% represents misclassifications, meaning where the model wrongly pronounces a particular image to belong to some other group that visually resembles. This could also arise from the intrinsic difficulty of the imagery task since it gets complicated when you encounter photos showing Chihuahua closely matching muffin in visual appearance or in cases where muffin is able to imitate visually that of a chihuahua.

An inspecting on misclassification shows that the model is slightly better classifying chihuahuas as muffins, where 12% of the mistakes come from. Errors arise when the appearance of the muffin resembles that of a chihuahua's fur, or when the color or texture of a chihuahua's ears seem similar to that of a muffin. On the other hand, we commit 3% of misclassifications mistaking muffins as chihuahuas. This tells us that the model is at a greater disadvantage when it comes to discerning certain characteristics from a photo of a chihuahua, especially when the hair or form of the dog appears similar to the uneven texture or shading appearance of a muffin.

The model is very successful when we show it a set of photos that are all high quality, well-lit and with the subject (a chihuahua or a muffin) aligned in the middle of the photograph. However, when we veer off from these ideal conditions, for multiple reasons such as when humans are partially covered in some photos, there is less light in the scene, or there are elements in the background, either noises or objects, which makes the algorithm second-guess its prediction, usually because there is an increase in misclassifications. The ideation tells us that the CNN model usually attains good performance. Further directions to improve this algorithm accuracy include how to make the model generalize well under difficult conditions.

1. **Comparision**

-Performance: Convolutional Neural Networks (CNNs) usually outperform conventional neural networks in image-related tasks due to theory ability to identify spatial hierarchies within the data. In the "Chihuahua vs. Muffin" task-CNN is where a fully connected neural network would encounter severe difficulties in the identification between the two distinct classes, given its flawed architecture in the complexity of visual features-that architects an accuracy of 85% for itself, while bleakly reducing normally below 70%, such is its plight. The traditional model views every pixel on an individual level, resulting in a flattened image that fails to recognize spatial interactions among pixels. Lack of convolutional layers in ordinary feedforward neural networks prevents such networks from identifying image features such as edges and textures-plain characteristics vital for image recognition-and so making classification tasks rather impossible et al.

CNNs: convolutional layers turn visual information of images into localized features such as textures, colors, and shapes, allowing them to perform far better when classifying images. Convolutional Neural Networks use pooling layers to curb the data dimensionality while enhancing several critical features-all things traditional neural networks can't do as well.

-Training Time: Though CNNs make impressive improvements, they cost exceedingly high in training time. Convolutional Neural Networks (CNNs) necessitate huge amounts of computation time because they use filters, employ convolutional layers, and can possess a large number of parameters. Training a convolutional neural network takes far more time and machine time than an ordinary neural network. A classical neural network is trained for instances, for 1 or 2 hours on a specific image data set; on the other hand, the convolutional neural networks may take from 2 up to 3 times longer due to the depth of the architecture and the volume of data.This causes convoluted procedures such as convolution, pooling, and backpropagation through several layers, all of which manifest a tremendous quantity of parameters to update.

Alternatively, the classical neural networks are simple so that they may be straightforward in training, and training would mainly involve only fully connected layers. . Their training time is often reduced greatly because they flatten the pixel input of the image and they do not have any of those computationally intensive processes that are characteristic of CNNs. This speed, however, comes at the expense of lower accuracy and effectiveness in image- or geographical-driven applications.

1. **Challenges and Solutions**

Through the laboratory session, numerous difficulties arose, particularly with model efficacy and assignment difficulties. The central challenge was to have a model that, without causing excessive disturbance, could nonetheless differentiate very similar things, chihuahuas, and muffins, varying subtly in feel and form, in their classification. Thus far, the first cut of the CNN ran into this problem, and accuracy was lowered with high probability of random parity.

In this respect, I focused on optimizing the CNN architecture by tweaking the filter dimensions and increasing the model's depth for more efficient feature extraction. Another way of augmenting the data could have been further training over many samples, by rotating, flipping, and zooming in-and-out, enabling the model to generalize better and control overfitting, hence allowing it to perform better in classifying new, unseen images.

On the other hand, the other thing to be tackled was increasing time for training CNNs because of the computation imposed on it. I therefore utilized GPU acceleration and improved the model training process with early stopping if performance reached a plateau, saving time in terms of computation. Such practices, in conjunction with the above-mentioned teenage qualities, allowed for very rapid iterations of model changing. The solutions were highly efficacious in increasing the effectiveness and capability of the CNN, hence allowing the successful execution of the classification task.

1. **Real-World Application**

The image classification approach used in this CNN for the "Chihuahua vs. Muffin" problem has a wide variety of possible applications in the real world. One area is medical imaging, CNNs may learn to detect diseases from X-rays, MRIs, or some other imaging technique, thereby assisting doctors in diagnosing disorders like cancers, fractures, or organ abnormalities. Other models are used in autonomous vehicles to identify people, traffic signs, and obstacles almost instantaneously, thereby ensuring safe navigation. Retail and e-commerce systems have been utilizing image classification methods for product classification and visual search, which allows consumers to simply upload a picture and find similar products. Finally, CNNs can be used in security and surveillance for face recognition and anomaly detection, giving the possibility of speeding up security processes by automatically detecting potential threats or illicit persons. This once again highlights, in other words, the general applicability of image classification techniques across various sectors.

1. **Ethical Considerations**

Ethical concerns are therefore paramount to consider in the design and use of image classification models, particularly those based on Convolutional Neural Networks (CNNs). A key worry is training data bias. If the dataset on which the model has been trained suffers from under-representation and reflects societal prejudices, the model is likely to perform poorly on some demographic groups. Colour-coded facial recognition algorithms have been found to have much higher error rates for people of colour and women. This can allow biased models to discriminate against people of colour and women, leading, for example, to unjust arrest or incorrect recruitment decisions (Buolamwini & Gebru, 2018).   
  
 Another concern is to privacy. Large datasets are often needed for image classification models, and some of them may contain data that is inherently private. The collection, storage and use of data needs to comply with data protection laws such as the General Data Protection Regulation (GDPR) in Europe, which require the data subject's clear and unambiguous consent. Failure to comply with the law can result in significant legal and reputational penalties. Models built using deep learning are often considered "black-boxes", because they can be extremely difficult to interpret by the user.   
  
 This lack of transparency can lead to suspicions and scurrilous infidelities, particularly in high-stakes applications such as healthcare, where a patient might want to know the reasons for a diagnosis (Lipton, 2016). It is hence quintessential that these models must be developed to foster trust and accountability; regarding ethical issues arising from image classification models concerning surveillance and security, those encompass civil liberty and the risk of abuse. Additionally, police use of facial recognition technology will bring about endless surveillance at the cost of seriously undermining individual rights, especially regarding those populations already subjected to unduly harsh monitoring-furthering their plight. Some of the fundamental dilemmas presented in a consideration of the utilization of such technology demand thorough consideration of their ethical implications, alongside the setting up of rules and regulations making up that very confined.   
  
 To sum up, despite the fact that image classification models-e.g., CNNs may be quite beneficial in several applications, their generation and use ought to arise before a well-considered stage. It is indispensable to ensure that any concerns beer-virus, privacy, transparency, and civil rights-have been clearly addressed in order to guarantee the responsible and ethical utilization of such technologies.

1. References

*Unsupervised Feature Learning and Deep Learning Tutorial*. (n.d.). <http://deeplearning.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/>

Zvornicanin, E. (2022, April 11). *Convolutional Neural Network vs. Regular Neural Network | Baeldung on Computer Science*. [Www.baeldung.com](http://Www.baeldung.com). <https://www.baeldung.com/cs/convolutional-vs-regular-nn>

‌Kumar, A. (2021, November 2). *Real-World Applications of Convolutional Neural Networks - Analytics Yogi*. Analytics Yogi.

<https://vitalflux.com/real-world-applications-of-convolutional-neural-networks/>