POZNAN UNIVERSITY OF TECHNOLOGY

PHD DISSERTATION - ABSTRACT

Restricted Boltzmann Machine as a binary image descriptors processor and its application in a mobile robot for scene recognition

Ograniczona maszyna Boltzmanna jako procesor binarnych deskryptorów obrazu oraz jej aplikacja w robocie mobilnym do celów przetwarzania wizji

Author:
Szymon Sobczak, MSc

Supervisors:
Dariusz Pazderski, PhD, DSc
Rafał Kapela, PhD

30/05/2023

Abstract

In recent years we have observed a dynamic growth of vision systems and artificial intelligence in science and many areas of industry. Due to computationally efficient resources, access to large quantities of data, and progress in science, it is possible to train very complex neural networks, that are capable of solving many image recognition problems, and whose effectiveness is comparable to human perception. However, the fact that solving complex classification problems often requires very large recognition systems has led to a focus on increasing the complexity of networks or enlarging their size, which in turn increases the complexity of recognition systems. Most of the currently used devices can handle those tasks, but it is not always possible to have access to efficient processing systems, which implies that large neural networks are not always applicable, especially in small devices, when price, size, and power consumption are the priorities. Another concern with complex neural architectures is that they require a large amount of training data, however accessing labelled training data is a frequent problem and may be expensive, time-consuming, and error-prone.

This dissertation presents a novel approach to visual data preprocessing that is focused on mitigating these difficulties. The proposed method introduces an additional stage of processing before the data is passed to a classification neural network. The classifier in this case may be a deep neural network which peforms an additional high-level feature extraction, or a shallow network that utilises the previously extracted features directly. This stage is composed of two layers, the first transforms an RGB image data to its binary feature representation, the second is a small and fast neural network that utilises the binary data to obtain its most important features, which can be then used for classification. The output of the second layer is a real-valued tensor representing feature vectors for each image pixel, that can be directly passed as input of a regular neural network. An additional novelty is an improved local binary descriptor, which embeds colour information in its feature vector, instead of having only a code of the shape of a given part of an image. The main advantage of this approach is that the neural layer can be trained in a fully unsupervised manner, also its processing time is short in comparison to widely known neural feature transformers. The experiments performed in this dissertation demonstrate that adding these layers may be successfully applied to increase the overall accuracy of a neural network or decrease the size of a neural network without losing its quality.

Furthermore, the preprocessing proposed in this study tackles other common image processing problems. The experiments show its robustness in terms of input data denoising, and likewise introduces a special metrics that

can be used for comparing images, or measuring the similarity between image datasets to test if a given neural network can be used for transform learning.

The mentioned techniques tend to decrease the overall complexity of neural structures, hence their potential use is in embedded devices. Robotics is one of the areas that utilise these and vision in robots is an important part of their control systems and the size, cost, and power consumption of processing units are significant factors in design. Therefore, a part of this study focuses on an application of the proposed method in a mobile robot equipped with a single camera and a relatively low efficient processing unit. Experiments performed with this application demonstrate that the previously presented and analysed method can be successfully applied in a real device, thus the suitability of using the preprocessing layer in embedded systems is experimentally proved.

 Szymon Szczęsny

Streszczenie

W ostatnich latach obserwujemy bardzo dynamiczny rozwój systemów wizyjnych i sztucznej inteligencji w różnych obszarach przemysłu, nauki i rozrywki. Wysoce wydajne zasoby sprzętowe, dostęp do dużej ilości danych oraz postęp badań w obszarze nowoczesnych technik identyfikacji i modelowania, przetwarzania danych i informatyki spowodowały, że możliwe jest stosowane bardzo złożonych sieci neuronowych w charakterze uniwersalnych aproksymatorów. W efekcie, takie architektury mogą rozwiązywać wiele problemów klasyfikacji obrazów, a ich efektywność jest zbliżona do ludzkiej percepcji. W ogólności, rozwiązywanie złożonych problemów klasyfikacji często wymaga bardzo skompilowanych systemów rozpoznawania. Badania w obszarze klasyfikacji obrazów skupiają się więc głównie na zwiększaniu złożoności sieci neuronowych lub bazują na wcześniej zdefiniowanych architekturach wymagających obliczeniowo. Większość obecnie używanych urządzeń może obsłużyć takie zadania, jednak dostęp do wydajnych jednostek obliczeniowych w niektórych przypadkach bywa ograniczony, przez co nie zawsze skomplikowane sieci neuronowe mogą być stosowane. Problem ten szczególnie dotyczy małych urządzeń gdzie ich rozmiar, cena i zużycie energii jest priorytetem. Dodatkowym problemem złożonych architektur jest to, że wymagają one bardzo dużych ilości oznaczonych danych trenujących, a dostęp do nich często jest trudny oraz może być kosztowny, czasochłonny i podatny na błędy.

Ta rozprawa doktorska przedstawia oryginalne podejście do wstępnego przetwarzania danych wizyjnych, którego celem jest zwiększenie efektywności rozpoznawania obrazów przy zachowaniu ograniczonego zapotrzebowania na moc obliczeniową. Zaproponowana metoda wprowadza dodatkową fazę przetwarzania bezpośrednio przed klasyfikującą siecią neuronową i proponuje etapy przetwarzania realizowane przez dwie warstwy. Pierwsza z nich przetwarza obraz RGB do jego reprezentacji w postaci wyekstrahowanych binarnych cech, druga jest niewielką siecią neuronową, która przetwarza binarne dane tak aby wydobyć z nich najważniejszych informacje i zależności między nimi, które można użyć dalej do klasyfikacji. Wyjście z drugiej warstwy jest tensorem wartości rzeczywistych, reprezentującą wektory cech dla każdego piksela. Tensor ten może być przekazany bezpośrednio do klasyfikującej sieci neuronowej. Klasyfikatorem w tym przypadku może być głęboka sieć neuronowa dodatkowo ekstrachująca cechy obrazu na wyższym poziomie abstrakcji lub płytka sieć neuornowa realizująca predyckję na wcześniej wydobytych cechach. Dodatkowym oryginalnym rozwiązywaniem jest ulepszony lokalny deskryptor binarny, który w swoim wektorze cech koduje nie tylko kształt danej część obrazu, ale również informacje o jego kolorze.

Główną zaletą takiego rozwiązywania, jest to że warstwa neuronowa jest trenowana w całkowicie nienadzorowany sposób, a także czas przetwarzania

przez nią danych jest niski w porównaniu do innych znanych neuronowych detektorów cech. Wyniki badań eksperymentalnych prowadzonych w ramach tej pracy wskazują, że dodanie proponowanych warstw przetwarzania może być z powodzeniem zastosowane w celu zwiększania jakości rozpoznawania przez sieci neuronowe, a także do zmniejszenia ich rozmiarów bez utraty jakości.

Zaproponowana metoda może być również użyta do rozwiązania innych powszechnie znanych problemów przetwarzania obrazu. Eksperymenty wskazują na celowość jej stosowania w przypadku odszumiania danych wejściowych. Wprowadza ona także specjalną metrykę, która może być zastosowana do porównywania obrazów, lub szacowania podobieństwa zestawów obrazów w celu testowania czy sieć neuronowa o danych parametrach może być użyta do rozwiązania innego problemu klasyfikacji bez konieczności dodatkowej optymalizacji wag.

Jednym z celów badań prowadzonych w ramach tej rozprawy jest dążenie do zmniejszenia ogólnej złożoności struktur neuronowych, przez co ich potencjalny obszar zastosowań to systemy wbudowane. Robotyka jest jedną z dziedzin która na nich bazuje, ponieważ wizja w systemach sterowania stanowi często ich istotną część. Ponadto czas, rozmiar i koszt jednostek obliczeniowych jest znaczącym czynnikiem w projektowaniu urządzeń przemysłowych z integrowanymi systemami rozpoznawania obrazu. W związku z tym, część pracy skupia się na aplikacji zaproponowanej metody w podsystemie percepcji kołowego robota mobilnego wyposażonego w jedną kamerę i relatywnie nisko wydajną jednostkę obliczeniową. Wykonane testy wskazują, że przedstawiony i przeanalizowany schemat przetwarzania danych wizyjnych może być skutecznie zastosowany w rzeczywistym urządzeniu. Pozwoliło to na eksperymentalne potwierdzenie słuszności użycia technik opracowanych w niniejszej rozprawie dla pewnej klasy systemów wbudowanych.

Szymon Szczęsny