dr hab. inż. Adam Milik, prof. PŚ.
Politechnika Śląska
Wydział Automatyki Elektroniki i Informatyki
Katedra Systemów Cyfrowych
ul. Akademicka 16
44-100 Gliwice

Gliwice, 10.12.2024

Recenzja rozprawy doktorskiej mgr. inż. Dominiki Przewłockiej-Rus dla Rady Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie

(podstawą opracowania recenzji jest uchwała Rady Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne z dnia 12.09.2024)

Tytuł rozprawy: **Metody kwantyzacji i akceleracji głębokich sieci neuronowych dla energooszczędnych systemów wizyjnych czasu rzeczywistego**

Tytuł rozprawy: **Deep Neural Networks Quantization and Acceleration Methods for Real Time Energy Efficient Vision Systems**

Autor rozprawy: mgr inż. Dominika Przewłocka-Rus

Promotor rozprawy: prof. dr hab. inż. Marek Gorgoń
Promotor pomocniczy: dr inż. Tomasz Kryjak

Dziedzina: Nauki Techniczne
Dyscyplina: Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne

1. Zagadnienia naukowe rozprawy – cel i teza pracy

Tematem rozprawy są metody kwantyzacji i akceleracji energooszczędnych głębokich sieci neuronowych dla systemów wizyjnych . Autorka w rozdziale 1 stawia następującą tezę: *Odpowiednie metody kwantyzacji parametrów sieci neuronowych pozwalają na znaczącą redukcję rozmiaru pamięci oraz złożoności obliczeniowej przy zachowaniu wysokiej dokładności oraz umożliwieniu implementacji wizyjnych systemów czasu rzeczywistego za pomocą platform sprzętowych o obniżonym poborze mocy.* (*Appropriate methods for quantising the parameters of neural networks allow a significant reduction in the memory and computational complexity of the models, while guaranteeing the preservation of a high accuracy and enabling implementation of real-time vision systems in low latency and low power hardware platforms.*)

A. Milik

Autorka przedstawia swoje osiągniecia za pomocą cyklu monotematycznego artykułów skupiających się wokół tezy podstawowej. W sposób niestandardowy w miejsce sformułowania tez szczegółowych autorka przechodzi do punktów podsumowujących jej wkład w przedstawione rozwiązania algorytmiczne i implementacyjne z zakresu budowy głębokich sieci neuronowych. Podsumowanie to należy uznać za przedstawienie wyników, których osiągniecie było celem pracy jednak nie zostało sformułowane wprost. Przytaczając za pracą następujące cele zdaniem autorki zostały osiągnięte:

1. Opracowanie metod uczenia sieci neuronowej, której współczynniki zostały przedstawione za pomocą zapisu wykładniczego, konstrukcję specjalizowanego sprzętowego układu mnożąco-akumulującego (MAC), połączenie warstwy obliczeń splotowych i normalizacji blokowej z uwzględnieniem specyfiki sieci (The proposal of methods for training neural networks quantised to weights of powers-of-two values, the design of a hardware architecture of special MAC operator (and the convolution layer), taking into account the particular form of such a network, and a method of fusion of the convolution and batch normalization layers, taking into account the particular form of such a network.)
2. Badanie wpływu różnych metod kodowania współczynników (linowe, logarytmiczne) na sieci neuronowe wykorzystywane w zaawansowanych systemach wizyjnych w celu określenia wpływu na dokładność obliczeń, złożoność obliczeniową oraz zapotrzebowania na zasoby pamięciowe a także efektywność energetyczną. (A series of experiments and analysis of the impact of different quantisation schemes (linear, logarithmic), with different target bit-widths, for neural networks used in advanced vision systems in embedded devices, to assess the impact on the accuracy, memory-computational complexity and energy efficiency of such systems.)

Uwzględniając przedstawioną tezę oraz szczegółowe kierunki prowadzonych badań stwierdzam, że praca doktorska mgr. inż. Dominiki Przewłockiej-Rus wpisuje się w aktualny nurt związany z projektowaniem głębokich sieci neuronowych ze szczególnym uwzględnieniem kwantyzacji współczynników, akceleracji obliczeń oraz redukcji poboru mocy. Prowadzone prace miały na celu uzyskanie sieci neuronowych przeznaczonych do analizy sygnałów wizyjnych. Platformą docelową były systemy przenośne. Charakteryzują się one ograniczoną liczbą zasobów sprzętowych a także oczekuje się od nich możliwie małego poboru energii.

A. Mirik

2. Zawartość i ocena merytoryczna

Recenzowana rozprawa doktorska składa się z rozszerzonego wstępu oraz cyklu dziewięciu artykułów przedstawiających dorobek naukowy autorki.

W rozdziale pierwszym autorka wprowadza czytelnika w tematykę pracy, ilustrując poruszane zagadnienia przeglądem literatury odnoszącym się do rozważanego problemu. Umiejscawia przedmiot swoich zainteresowań w obrębie zagadnień projektowania sieci neuronowych, kładąc szczególny nacisk na redukcję złożoności obliczeniowej oraz pobór energii przez układ obliczeniowy. Jak wykazuje, niezależenie od docelowej platformy implementacji redukcja liczby współczynników sieci wpłynie korzystnie na zapotrzebowanie na pamięć współczynników oraz pobór energii. Autorka wskazuje, że najlepsze efekty uzyskuje się za pomocą skoordynowanej metody projektowania algorytmu oraz sprzętu do jego realizacji.

Rozdział drugi w sposób syntetyczny przedstawia rozwiązania i osiągniecia. Szczególnym wkładem jest redukcja rozmiaru bitowego współczynników co przekłada się na redukcję zapotrzebowania na pamięć niezbędną do ich przechowywania oraz złożoność obliczeniową. Zastosowanie wykładniczego zapisu współczynników pozwala na znaczną redukcję liczby bitów, na których jest przechowywany współczynnik oraz radykalne uproszczenie struktury sprzętowej jednostki mnożąco-akumulującej. W przypadku dedykowanej implementacji sprzętowej w układach programowalnych FPGA uzyskuje się redukcję przestrzeni pamięci koniecznej do przechowywania współczynników. Redukcja rozmiaru bitowego współczynników powoduje również redukcję złożoności sprzętowej układów mnożących. Pozytywnym efektem przeprowadzonej minimalizacji jest redukcja mocy pobieranej przez układ obliczeniowy.

Wykładnicza metoda kwantyzacji współczynników stanowi bardzo interesujące i cenny element w zakresie implementacji sieci neuronowych. Należy podkreślić, że wykorzystanie zapisu wykładniczego pozwala na istotne zwiększenie zakresu reprezentowanych wartości przez liczbę zapisaną na zadanej liczbie bitów. Należy zauważyć również, że organizmy żywe odbierając bodźce fizyczne posiadają percepcję o charakterze wykładniczym. Zaproponowany sposób zapisu wartości ogranicza mantysę do jednej pozycji. Przyjmując koncepcję standardu IEEE754 reprezentacja liczby zostaje ograniczona do wykładnika i znaku. W celu przeprowadzenia obliczeń konieczne jest określenie kodowania, dla którego liczba przyjmuje wartość 0 jako odpowiedniego kodowania liczby w zapisie wykładniczym.

Autorka wraz ze sposobem kwantyzacji współczynników zaproponowała metodę uczenia sieci polegającą na propagacji wstecznej z wykorzystaniem współczynników zmiennoprzecinkowych, które w następnym kroku procesu podlegają kwantyzacji wykładniczej. Schematycznie odwzorowanie jednostki mnożąco-akumulującej z kwantyzacją wykładniczą

A. Mick

znajdujemy na rysunku 2.1. Przedstawiony model przepływu danych nie uwzględnia specyfiki odwzorowania sprzętowego. Autorka w swych pracach korzysta z układów FPGA rodziny Virtex 7 lub Artix 7 (Xilinx/AMD). Implementując wspomniany układ w zasobach logicznych ogólnego przeznaczenia, nie znajdujemy dokładnych rozważań dotyczących złożoności implementacji oraz wydajności obliczeniowej. Należy postawić pytanie o koszt odwzorowania układu przesuwającego pełniącego role multiplikatora dla współczynników wykładniczych. W rozważanym przypadku jest to implementacja struktury złożonej z multiplekserów odwzorowanych za pomocą bloków tablicowych oraz pomocniczych multiplekserów plastra logicznego. Z kolei w przypadku akumulatora pewne zaniepokojenie budzi wykorzystanie oddzielnego układu dopełnienia do wyznaczenia wartości liczby przeciwnej. Warto zauważyć, że niezwykle łatwo można połączyć wyznaczenie liczby przeciwnej w zapisie uzupełnienia do 2 z sumatorem akumulatora. Podsumowując możliwe jest uzyskanie regularnej kompaktowej struktury, co z kolei jest istotne w przypadku odwzorowania w układzie programowalnym FPGA. Prezentacja wyników implementacji w tabeli 2.3 została pokazana bez kontekstu implementacyjnego. Czytelnik nie ma możliwości oceny skuteczności implementacji bez podania precyzyjnych parametrów implementowanej sieci. Odniesienie się do dostępnych elementów w układzie pozwala jedynie wskazać czy implementacja będzie możliwa w pojedynczym układzie scalonym. Należy wskazać jeszcze jeden istotny element pominięty w opisie. Jest nim harmonogram obliczeń wraz z układem sterowania przepływem danych w jednostce obliczeniowej. Przedstawienie tych elementów jest istotne z punktu widzenia efektywności wykorzystania zasobów obliczeniowych.

Ważnym osiągnieciem implementacyjnym jest opracowanie akceleratora sprzętowego sieci XNOR. Przeznaczona jest on do rozpoznawania znaków drogowych Do prowadzenia obliczeń wykorzystano operacje bitowe oraz zliczanie bitów o wartości 1 w słowie wyjściowym. Podczas prac nad układem akceleratora sprzętowego zostaje wydany program FINN, który jest przeznaczony do generacji sprzętowej struktury sieci neuronowej. Uzyskane wyniki porównano z wynikami uzyskanymi za pomocą programu generatora struktur sieci neuronowych FINN. Przedstawiona implementacja uzyskana za pomocą programu FINN dedykowana jest dla układów rodziny ZYNQ, będących połączeniem systemu mikroprocesorowego z platformą FPGA. Akcelerator sieci neuronowej został włączony w programowy łańcuch przetwarzania danych. Nie określono jednak czy układ jest całkowicie pasywny, czy też jest zdolny do autonomicznego pobierania danych z bufora w pamięci.

Rozdział drugi podsumowuje syntetyczny zestaw osiągnięć autorki w zakresie projektowania i implementacji sieci neuronowych.

Rozdział trzeci jest zbiorem wybranych publikacji autorki mających na celu dowiedzenie postawionej tezy. Składa się z dziewięciu artykułów, które po krótce zostaną scharakteryzowane.

A. Mirk

1. XNOR CNNs in FPGA: Real-time Detection and Classification of Traffic Signs in 4K – a demo

W artykule przedstawiono wstępne badania nad akceleracją sprzętową sieci XNOR przeznaczoną do detekcji znaków drogowych. Istotną trudnością w ocenie zaprezentowanych rozwiązań jest skrócona forma artykułu mieszcząca się w całości na jednej stronie (rozszerzone streszczenie). Brak tutaj posłowia autorki lub też zmieszczenia fragmentów plakatu prezentowanego na konferencji. Istotą w takim przypadku jest sposób implementacji i wykorzystania zasobów. Wnioskując z opisu wykorzystano techniki odwzorowania zachłannego z wykorzystaniem harmonogramowania ASAP lub ALAP. Zastosowane techniki mogą mieć charakter niejawny wynikający z bezpośredniego odwzorowania sprzętowego.

2. Optimisation of a Siamese Neural Network for Real-Time Energy Efficient Object Tracking

Artykuł przedstawia implementację neuronowej sieci syjamskiej przeznaczonej do śledzenia obiektów. Istotą pracy jest zbadanie wpływu kodowania współczynników sieci na rozmiar pamięci niezbędnej do przechowywania współczynników przy zachowaniu własności detekcyjnych. Przedstawione rozważania mają istotne znaczenie w przypadku implementacji z wykorzystaniem dedykowanej platformy sprzętowej. W przypadku wykorzystania platformy mikroprocesorowej efektem jest redukcja rozmiaru pamięci współczynników. Istotnym mankamentem jest brak opisu charakterystyki sieci co utrudnia szacowanie zasobów pamięciowych przez czytelnika. Autorzy również zwracają uwagę na zapotrzebowanie energetyczne układu. Redukcja rozmiaru współczynników sieci pozwala na redukcję zasobów koniecznych do przeprowadzenia obliczeń. Autorzy nie wskazują na wykorzystanie rozszerzonych technik kontroli poboru mocy bazujących np. na selektywnym taktowaniu aktywnych elementów systemu obliczeniowego. Ze względu na brak szczegółów implementacji sprzętowej trudno się odnieść w sposób pełny do redukcji pobieranej energii przez jednostkę obliczeniową.

3. Quantised Siamese Tracker for 4K/UltraHD Video Stream – a demo

Artykuł w formie rozszerzonego streszczenia przedstawia zarys implementacji neuronowej sieci syjamskiej przeznaczonej do śledzenia obiektów w obrazie wysokiej rozdzielczości (3840 x 2160). Istotą rozważania jest wyznaczenie współczynników o możliwie niewielkich długościach bitowych dla poszczególnych warstw. Artykuł nie przedstawia szczegółów opracowanej architektury jednostki obliczeniowej oraz wyników implementacji.

4. Exploration of Hardware Acceleration Methods for an XNOR Traffic Signs Classifier

Artykuł przedstawia metodę odwzorowania sieci neuronowej typu XNOR przeznaczonej do klasyfikacji znaków drogowych. Praca stanowi rozwinięcie artykułu 1, w którym wstępnie zasygnalizowano opracowanie akceleratora sprzętowego dla sieci typu XNOR. Opracowany

A. Mirk

akcelerator stanowi ciekawe ujęcie konstrukcji opracowanej przez autorów. Przedstawiony opis wskazuje na rozdzielenie bloków poszczególnych warstw. Należy wskazać na bardzo interesujące podejście w zakresie potencjalnego potokowego wykonania operacji. Przedstawione podsumowanie wykorzystania zasobów sprzętowych nie zostało skomentowane. Szczególnie istotne jest dokonanie klasyfikacji zasobów pod kątem elementów przechowujących współczynniki sieci do elementów wykonujących obliczenia i sterujących procesem obliczeniowym. Uzyskane wyniki porównano ze strukturą uzyskaną za pomocą programu FINN przeznaczonego do generacji sprzętowych struktur sieci neuronowych. Wyniki uzyskane za pomocą programu FINN w zestawieniu z wynikami opracowanej przez autorów struktury są dość zaskakujące. Istotną różnicą jest wykorzystanie elementów logicznych. Szczególnym zaskoczeniem jest całkowite wyeliminowanie bloków DSP48 w rozwiązaniu finalnym. Autorzy zwracają również uwagę na pobór energii przez sieć. W zestawieniu wyników podają moc pobieraną przez układ wraz z liczbą przetworzonych ramek obrazu w czasie jednej sekundy. Wydaje się, że bardziej obrazowym podejściem byłoby wyznaczenie energii potrzebnej na przetworzenie pojedynczej ramki. Taki wskaźnik miałby charakter normatywny. W przypadku układów programowalnych można zastosować techniki redukcji mocy polegające na selektywnym taktowaniu jednostek obliczeniowych aktywowanych wraz z przepływem przetwarzanych danych. W przypadku rozwiązań o dużej wydajności obliczeniowej można wykorzystywać sposób działania polegający na przetwarzaniu danych na żądanie, co skutecznie redukuje moc średnią pobieraną przez urządzenie.

5. Power-of-Two Quantization for Low Bitwidth and Hardware Compliant Neural Networks

Prace nad redukcją złożoności obliczeniowej sieci neuronowych skupiają się nad sposobem przedstawienia współczynników w taki sposób, aby zapisać je na możliwie najmniejszej liczbie bitów zachowując efektywny proces klasyfikacji. Oznacza to zmianę podejścia do sposobu reprezentacji wartości współczynników wagowych. Zapis pozycyjny wagowy pozwala na dokładne przedstawienie liczb całkowitych, natomiast jego istotną wadą jest niewielki zakres reprezentowanych wartości. W konsekwencji, aby przedstawić niezbędny zakres wartości należy wykorzystać liczby o znacznej długości. Dodatkowo można wskazać na względnie szeroki przedział wartości 0. Z punktu widzenia dynamiki wartości, zapis wykładniczy pozwala na przedstawienie znaczne większego przedziału wartości aniżeli w przypadku liczb całkowitych zapisanych na tej samej liczbie bitów. W omawianym artykule skupiono się na trzech aspektach związanych ze zmianą sposobu reprezentacji współczynników sieci. Istotą rozwiązania jest wykorzystanie zapisu wykładniczego współczynników sieci. Przedstawione rozwiązanie uwzględnienia sposób reprezentacji wartości współczynników w procesie uczenia, tak aby odpowiednio dobrać ich wartości. Zmniejszenie rozmiaru bitowego współczynników przekłada się na redukcję rozmiaru pamięci koniecznej do ich przechowywania. Sposób reprezentacji współczynników determinuje również sposób realizacji obliczeń. Współczynniki w zapisie pozycyjnym wymagają implementacji układu mnożącego,

A. Mick

natomiast zapis wykładniczy ogranicza się do skalowania argumentu, co przekłada się na przesunięcie o wskazaną przez wykładnik liczbę pozycji. Złożoność układu przesuwającego oraz czas propagacji argumentów są znacznie mniejsze aniżeli w przypadku układu mnożącego. Zapotrzebowanie na zasoby zostało przedstawione w tabeli 10. Dodatkowo autorzy zwracają uwagę, że wraz z redukcja złożoności układowej następuje redukcja mocy pobieranej przez układ

6. Towards real-time and energy efficient Siamese tracking – a hardware-software approach

Artykuł przedstawia kontynuację rozważań nad neuronową siecią syjamska przeznaczoną do śledzenia obiektów. Autorzy stawiają sobie za cel implementację efektywnego systemu śledzenia wizyjnego dla systemów przenośnych. W odróżnieniu od rozwiązań stacjonarnych wykorzystanie energochłonnych akceleratorów graficznych nie jest możliwe. Autorzy kierują swoje zainteresowania na implementację mieszaną sprzętowo-programową. W zaproponowanym rozwiązaniu system mikroprocesorowy przygotowuje dane dla sprzętowego akceleratora obliczeniowego sieci neuronowej. Implementacja akceleratora odbywała się z wykorzystaniem narzędzia FINN. W pracy nie został sprecyzowany sposób wymiany informacji z akceleratorem. W celu wyszukania najefektywniejszego rozwiązania dokonano przeszukania przestrzeni rozwiązań. Przeprowadzono to poprzez zmianę parametrów generowanych jednostek obliczeniowych za pomocą narzędzia FINN. W wyniku powstało zestawienie porównujące wydajność obliczeniową oraz pobór mocy. Autorzy nie wprowadzili współczynnika normowanego pozwalającego na określenie uśrednionego zużycia energii koniecznej do przetworzenia ramki obrazu. W przypadku urządzeń przenośnych pozwalałoby to na łatwe porównanie efektywności energetycznej. W tabeli 4 zebrano wyniki opracowanych rozwiązań. Zaskakującym jest wykorzystanie aż 104.85% dostępnych bloków tablicowych LUT. Autorzy dokonują podziału na LUT i LUTRAM. Oba elementy należą do tej samej grupy funkcjonalnej bloków tablicowych, różnicą jest sposób wykorzystania jako układy logiczne (LUT) lub pamięci o dostępie swobodnym (LUTRAM). W tabeli 5 przedstawiono ciekawą analizę rozkładu czasu obliczeń pomiędzy elementami systemu. Warto zauważyć, że rozdzielenie procesów obliczeniowych powinno umożliwić pracę z przeplotem co podniosłoby sprawność działania układu.

7. Energy Efficient Hardware Acceleration of Neural Networks with Power-of-Two Quantisation

W tym artykule autorzy szczególny nacisk kładą na przedstawienie wpływu kodowania na zużycie energii. Rozważania prowadzoną są dla uprzednio przedstawionej metody kodowania wykładniczego (artykuł 3). Uzyskanie znacznej dynamiki wartości współczynników za pomocą reprezentacji wykładniczej pozwala na istotne zmniejszenie liczby bitów przy zachowaniu poprawnych własności przetwarzania sieci. Istotą jest zaprojektowanie układów mnożąco akumulujących o niewielkiej złożoności sprzętowej. Pierwszym z rozważanych rozwiązań jest

A. Mitik

klasyczny układ z multiplikatorem pokazany na rysunku 2. Przedstawione rozwiązanie dokonuje bezpośrednio mnożenia wagi zapisanej w sposób wykładniczy. Autorzy nie uwzględnili konieczności przekodowania wagi z systemu wykładniczego do pozycyjnego. Równocześnie multiplekser znajdujący się za układem mnożącym jest zbędny w przypadku umieszczenia transkodera. Rozwiązanie przedstawione na kolejnym rysunku odnosi się do reprezentacji wykładniczej wykorzystując przesuniecie bitowe (układ przesuwający). Nieco zaskakujące jest sterowanie wymuszeniem wartości 0. Iloczyn logiczny sygnału ZERO_WEIGHT oraz wektora współczynnika w_in nie jest konieczny. Sygnał ZERO_WEIGHT wystarcza do wymuszenia wartości 0. Niejasne jest również powód wyodrębnienia operacji dopełnienia (zmiany znaku liczby). Schemat blokowy sugeruje wykorzystanie dodatkowych elementów, co nie jest konieczne. Implementacja zmiany znaku powinna zostać wykonana przez negację argumentu akumulatora oraz ustawienie przeniesienia wejściowego sumatora akumulacyjnego. Wskazanie na multiplekser wymuszający wartość 0 jako element ograniczający pobór energii również jest niepełne. Należy zwrócić uwagę, że nie następuje stłumienie działania układu mnożącego. Istotą ograniczenia poboru mocy jest eliminacja propagacji zmian w obrębie układu. Ostatecznie autorzy wykazują redukcję zasobów logicznych niezbędnych do implementacji układu obliczeniowego przy wykorzystaniu techniki kodowania wykładniczego współczynników. Uproszczenie struktury pozwala również osiągnąć redukcję pobieranej mocy. Zamieszczone w tabeli 2 wyniki poboru mocy przez układ wymagają rozszerzonej analizy. W wyniku optymalizacji i zastosowania kodowania wykładniczego zmniejszono liczbę bloków tablicowych do 80,8% przerzutników do 73,2% oraz moc niezbędna do zasilania została zredukowana do 70,6%. Uzyskano proporcjonalnie większą redukcję mocy aniżeli wskazuje to liczba elementów. W przypadku przerzutników energia rozpraszana przez wejścia sygnału zegarowego powinna kształtować się na poziome ich redukcji. Redukcja pobieranej mocy jest możliwa przez ograniczenie aktywności przełączeniowej części kombinacyjnej. W tym przypadku konieczne jest przeprowadzenie analizy przełączeń za pomocą odpowiednio zinstrumentalizowanego modelu pozwalającego na rejestrację aktywności przełączeniowej poszczególnych elementów po implementacji układowej. Wyniki wskazują na istotne ograniczenie liczby przełączeń po zmianie sposobu kodowania współczynników wagowych.

8. Power-of-Two Quantized YOLO Network for Pedestrian Detection with Dynamic Vision Sensor

W artykule przedstawiono oryginalne opracowanie implementacji sieci YOLOv5s do wykrywania przechodniów. W celu uzyskania sieci przeznaczonej do implementacji w układach programowalnych o zredukowanym poborze mocy przeprowadzono badanie możliwości zastosowania współczynników z kodowaniem wykładniczym. Na podstawie przeprowadzonych eksperymentów autorzy wykazali, że zastosowanie kodowania wykładniczego współczynników pogarsza nieznacznie skuteczność sieci. Odnotowany spadek

A. Mikik

wykrywalności w stosunku do współczynników wagowych wyrażonych za pomocą liczb zmiennoprzecinkowych nie przekraczał 3,5%. Istotnym elementem opracowania był sposób uczenia sieci oraz metod wprowadzania kwantyzacji współczynników sieci. Przedstawione efekty wskazują na możliwość zastosowania wyników badań w energooszczędnych systemach wbudowanych

9. PowerYOLO: Mixed Precision Model for Hardware Efficient Object Detection with Event Data

Celem autorów jest opracowanie sieci neuronowej wykrywającej pieszych oraz inne obiekty drogowe dedykowanej do zastosowań w przemyśle motoryzacyjnym. Specyfika zastosowania wymaga użycia systemów wbudowanych o ograniczonej liczbie zasobów oraz silnie ograniczonym budżecie mocy dysponowanej do zasilania urządzenia. Artykuł jest kontynuacją prac nad sieciami YOLO oraz zastosowaniem dynamicznych czujników wizyjnych (Dynamic Vision Sensors – DVS). Jak zaznaczają autorzy zaleta czujnika DVS jest niższy pobór mocy oraz szeroki zakres czułości. W zaproponowanym rozwiązaniu sieci YOLO w części obliczeń splotowych zmieniono sposób kwantowania współczynników z liniowego na wykładniczy. Istotą zamiany metod kwantyzacji jest ograniczenie rozmiaru współczynników oraz zastąpienie operacji mnożenia operacją przesunięcia. W konsekwencji uzyskuje się podwójna korzyść objawiającą się redukcją rozmiaru pamięci współczynników a jednocześnie uproszczeniem architektury jednostki obliczeniowej. W zespole funkcji aktywacji zastosowano współczynniki kodowane liniowo na 8 bitach. Ostatecznie powstaje konstrukcja sieci o współczynnikach wykładniczych dla warstwy splotowej oraz liniowych dla warstwy aktywacji. W wyniku zaproponowanych modyfikacji możliwe było zredukowanie rozmiaru sieci 8 krotnie w porównaniu do implementacji ze współczynnikami kodowanymi za pomocą liczb zmiennoprzecinkowych. Uzyskanie powyższych efektów było możliwe dzięki zastosowaniu skoordynowanego projektowania obejmującego pełny system złożony z urządzenia obrazującego, algorytmu oraz platformy realizującej obliczenia.

Podsumowując analizę artykułów należy odnotować bardzo wysoki udział procentowy autorki w poszczególnych pozycjach. W pewnych przypadkach udział jest zaskakująco wysoki (np. artykuł 4) w odniesieniu do liczby autorów uczestniczących w powstaniu pracy. Dodatkowo w treści odnajdujemy podziękowania dla osób niewymienionych jako autorzy publikacji a mających udział w powstaniu układu sprzętowego akceleratora. We wszystkich artykułach poza artykułami w formie rozszerzonego streszczenia [1,4] odnajdujemy odpowiednio dobrany zestaw źródeł, na które powołują się autorzy. Do części opisowej autorka dołącza również wykaz 23 (w tym 9 publikacji z udziałem własnym zamieszczonych w dalszej części), na które powołała się w rozszerzonym wstępie do artykułów. Przedstawione przez doktorantkę publikacje doczekały się wielu cytowań, co wskazuje na zainteresowanie środowiska naukowego proponowaną tematyką oraz docenienie przedstawionych rozwiązań.

A. Mitik

Na podstawie lektury pracy można stwierdzić, że postawiona w rozdziale 1 teza została udowodnione a cele osiągnięte przez autorkę. W szczególności autorka przedstawiła oryginalne opracowania następujących zagadnień:

- Metod uczenia dla sieci neuronowych wykorzystujących współczynniki kodowane za pomocą zapisu wykładniczego na niewielkiej liczbie bitów (4 bity), których skuteczność dorównuje modelom sieci wykorzystującym współczynniki zmiennoprzecinkowe
- Opracowanie sprzętowego układu mnożąco-akumulującego dla sieci wykorzystujących współczynniki w zapisie wykładniczym. Zastosowanie współczynników w zapisie wykładniczym umożliwia zredukowanie zapotrzebowania na zasoby sprzętowe i pamięciowe oraz znaczące (około 2 krotne) zredukowanie poboru mocy przez urządzenie
- Opracowanie architektury sprzętowej dla splotowej sieci neuronowej wykorzysującej współczynniki o zapisie wykładniczym charakteryzującej się zredukowanym zapotrzebowaniem na zasoby sprzętowe, wyższą częstotliwością maksymalną pracy oraz zredukowanym poborem energii w stosunku do implementacji dla współczynników kodowanych liniowo
- Normalizacja wyników uzyskanych w warstwie splotowej uwzględniającej specyfikę sieci z kodowaniem wykładniczym współczynników
- Zestaw eksperymentów badających wpływ różnych sposobów kodowania współczynników sieci neuronowej (liniowe, logarytmiczne) stosowanych w zaawansowanych systemach wizyjnych w celu określenia wpływu na dokładność obliczeń, złożoność obliczeniową, zapotrzebowanie na zasoby pamięciowe oraz efektywność energetyczną
- Opracowanie architektury sprzętowo-programowej sieci neuronowej dla systemu wizyjnego umożliwiającego rozpoznawanie znaków drogowych, śledzenie obiektów oraz wykrywanie pojazdów i pieszych

3. Uwagi krytyczne, pytania

Podczas obrony publicznej chciałbym poznać zdanie doktorantki na poniższe zagadnienia:

1. Przedstawione w artykule 7 rysunki nie pozwalają na jednoznaczną interpretację zastosowanego odwzorowania sprzętowego. Na podstawie specyfikacji wykorzystanych zasobów można wnioskować, że zostały wykorzystane techniki obliczeń potokowych o czym świadczy liczba przerzutników w stosunku do bloków tablicowych LUT. Jak wygląda odwzorowanie technologiczne jednostki mnożąco-akumulującej będącej podstawowym elementem sieci neuronowej przedstawionej w artykule?

A. Milik

2. W wielu miejscach autorka określa, że zaproponowane rozwiązania pobierają mniej energii. Szczególnie wyczuwalne staje się to w przypadku rozwiązań wykorzystujących kodowanie wykładnicze współczynników wagowych. Wydaje się celowym określenie energii niezbędnej do przeprowadzenia obliczeń. Jak można dokonać takiej oceny dysponując precyzyjnym odwzorowaniem komórki podstawowej? Jakie inne metody ograniczenia pobieranej energii przez układ można dodatkowo zaproponować w przypadku istotnego zapasu mocy obliczeniowej w stosunku do wymagań stawianych przez zadanie?

3. Niezwykle interesująca jest konstrukcja akceleratora sprzętowo-programowego, która zasługuje na szersze przedstawienie szczegółów implementacji. Wykorzystany układ ZYNQ Ultrascale+ posiada niezwykle interesującą architekturę wykorzystującą magistralę AXI do powiązania elementów systemu. W przedstawionym projekcie nie zostało jasno sprecyzowane czy akcelerator jest całkowicie pasywnym elementem systemu czy też posiada możliwość aktywnej wymiany danych. Jak rozwiązano przyłączenie akceleratora do systemu za pomocą magistral AXI? Jaki harmonogram wykonania obliczeń został zastosowany? Czy zaimplementowane obliczenia są realizowane w sposób potokowy?

4. W pracach można odczuć pewien brak odniesienia się do sposobu realizacji obliczeń przez akceleratory sprzętowe sieci neuronowych. W jaki sposób zostały rozmieszczone operacje i zorganizowana wymiana informacji z pamięciami? Jakie metody zostały zastosowane do akceleracji obliczeń oraz zwiększenia częstotliwości sygnału zegarowego?

4. Uwagi szczegółowe

Przedstawiona rozprawa ma również swoje słabe strony i pewne niedociągnięcia. Chciałbym podkreślić, że znaczna część uwag ma charakter polemiczny.

W streszczeniu akapit 2 autorka formułuje następujące stwierdzenie: „requires an approach characterised by **consistency** between the algorithmic solution and the target platform on which it will run" W tym przypadku zaskakujące jest stwierdzenie o wymaganej spójności pomiędzy modelem a jego odwzorowaniem technologicznym. W przypadku odwzorowania technologicznego oczekuje się, że nastąpi wierne odwzorowanie modelu w strukturze sprzętowej. Przez wierne odwzorowanie należy rozumieć zachowanie wyników przetwarzania. Dopuszczenie do niezachowania powyższego wymagania prowadzi do całkowitej dowolności funkcjonalnej odwzorowania.

W streszczenie wzór 2.2 przedstawiający normalizację wyników uzyskanych w warstwie splotowej sieci nie wyjaśnia znaczenia wszystkich zmiennych. W formalnej pracy autorka nie powinna dopuszczać do pojawienia się niejasności oznaczeń.

A. Milik

W artykule 1 sekcja III.A: „Implementation of a fully connected layer assumes the use of a larger BRAM memory" Pamięci blokowe RAMB są elementami o określonej architekturze, w konsekwencji nieuprawnionym jest stwierdzenie o użyciu większej pamięci RAMB. W tym przypadku możemy się domyślać, że istotą jest zwiększenie zapotrzebowania na rozmiar bufora, który należy odpowiednio skonstruować z istniejących zasobów RAMB.

W artykule 4 sekcja 4.1 (str. 51) „The filter weights are stored in the on-board registers, therefore the utilisation of BRAM memory is reduced, and the time needed to read the weights is shortened." Przytoczone stwierdzenie nieprecyzyjnie wskazuje sposób przechowania współczynników stałych. Ze względu na brak możliwości odniesienia się do schematów blokowych lub fragmentów opisu należy wnioskować, że autorzy co najmniej niewłaściwie określili sposób przechowania współczynników. Próba wykorzystania rejestrów ogólnego przeznaczenia do przechowywania stałych jest metodą nieefektywną. Obrazowo opisując możliwe jest zapisanie do 8 bitów w każdym plastrze logicznym. Oznacza to, że współczynniki w zależności od sposobu opisu zostały odwzorowane za pomocą bloków tablicowych LUT występujących w plastrach logicznych typu L lub też rozproszonych pamięci zbudowanych z bloków tablicowych LUT występujących w plastrach logicznych typu M. W tym przypadku uzyskuje się ponad 32 krotnie większą gęstość upakowania informacji w plastrze logicznym (4 x 64 bity).

W artykule 6 W tabeli 4 (str. 76) zebrano wyniki opracowanych rozwiązań. Autorzy dokonują podziału na LUT i LUTRAM. Oba elementy należą do tej samej grupy funkcjonalnej bloków tablicowych przy czym możliwość modyfikacji do pamięci rozproszonej posiadają bloki tablicowe tylko w plastrach typu M. Zaskakującym jest wykorzystanie 104.85% dostępnych bloków tablicowych LUT dla wersji układu V6.

W artykule 7 sekcja 4 (str. 87) rysunek 2 przedstawia układ mnożący przy czym współczynnik wagowy jest przedstawiony w zapisie wykładniczym dodatkowo uzupełnionym o wartość 0. W konsekwencji przedstawiony układ powinien zawierać blok przekształcający współczynnik wagowy w zapisie wykładniczym na wartość w zapisie pozycyjnym. Ponadto sterowanie multipleksera o dwóch wejściach danych za pomocą wielobitowego słowa wymaga również odpowiedniego przekształcenia (przekodowania) słowa do wartości dwustanowej.

A. Mlik

5. Ocena końcowa rozprawy

Uważam, że przedstawione w rozprawie doktorskiej oryginalne metody oraz rozwiązania sprzętowe, wnoszą istotny wkład w rozwój dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne. Do szczególnych osiągnięć autorki należy zaliczyć opracowanie akceleratorów sprzętowych sieci neuronowych o niskim zapotrzebowaniu na zasoby sprzętowe oraz zredukowanym poborze mocy ze szczególnym uwzględnieniem kodowania wag za pomocą systemu wykładniczego, opracowanie metod uczenia dla kodowania wykładniczego oraz niezbędnych testów porównawczych. Wyniki swoich badań autorka przedstawiła w artykułach publikowanych w renomowanych czasopismach, na konferencjach zawiązanych z sieciami neuronowymi i zastosowaniami układów programowalnych.

Podsumowując stwierdzam, że rozprawa doktorska mgr. inż. Dominiki Przewłockiej-Rus spełnia wymagania stawiane w ustawie „Prawo o szkolnictwie wyższym i nauce", zatem wnoszę o dopuszczenie rozprawy doktorskiej do publicznej obrony i dalszych etapów przewodu doktorskiego w dyscyplinie Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne.

Adam Milik