

dr inż. Witold Czajewski
(tytuł, imię i nazwisko kierującego pracą)

Opinia pracy dyplomowej inżynierskiej

Dyplomant: **Piotr Heinzelman** Kierunek: **Informatyka stosowana**

Specjalność: **Informatyka stosowana**

Tytuł pracy: **Porównanie wydajności wybranych języków programowania w realizacji sieci neuronowych do przetwarzania obrazów**

Ocena: **3.0**

Treść opinii

1. Zgodność tytułu pracy dyplomowej z jej treścią.

Tytuł pracy jest zasadniczo zgodny z jej treścią. Praca dotyczy porównania czasów działania wybranych algorytmów związanych z sieciami neuronowymi w różnych środowiskach/językach dla kilku klas zadań, od regresji liniowej po detekcję obiektów. Jednocześnie należy zauważyć, że w wielu miejscach w praktyce porównywane są nie tyle „języki programowania”, co gotowe ekosystemy/biblioteki.

2. Wartość merytoryczna pracy (identyfikacja problemu, sformułowanie celu, dobór i sposób wykorzystania narzędzi, rozwiązanie zadania badawczego / projektowego / technologicznego / organizacyjnego)

Autor formułuje cel pracy w sposób jasny: identyfikacja podejść minimalizujących czas treningu i wnioskowania w zadaniach przetwarzania obrazów oraz porównanie wydajności wybranych środowisk w zestawie zadań (regresja, MLP, CNN, „głębokie” CNN, prosty detektor, YOLO). Zakres eksperymentów, przynajmniej we wstępnej deklaracji, jest stosunkowo szeroki, a czytelnik dostaje wiele pomiarów czasów działania. Jednak zasadniczy problem merytoryczny dotyczy porównywalności badań i jakości wnioskowania na podstawie zebranych danych. W wielu miejscach „język” jest tylko cienką warstwą uruchomieniową dla biblioteki, która wykonuje właściwą pracę (często na GPU), a w innych miejscach implementacja jest autorska albo półautorska. Przykładowo: Autor pisze własny klasyfikator MLP w Javie, ale porównuje go z gotowymi rozwiązaniami bibliotecznymi w innych językach — co z góry ustawia nierówne warunki (inny poziom optymalizacji, inne wykorzystanie sprzętu, inna dojrzałość kodu). Podobnie w części dot. C++/CUDA Autor wprost wskazuje, że wykorzystywany kod pochodzi od innych autorów i nie do końca wiadomo, co w nim jest. Taki dobór materiału obniża „inżynierski” wkład własny oraz utrudnia ocenę, czy różnice wynikają z języka, biblioteki, czy jakości implementacji.

Autor deklaruje, że kluczowym kryterium porównywania badanych rozwiązań jest ich szybkość, a analiza jakości (np. dokładności) została świadomie pominięta, tymczasem Autor w kilku miejscach pokazuje wyniki analizy jakościowej dla niektórych przypadków. Takie podejście jest niespójne.

Zastrzeżenia budzi też sposób przeprowadzenia eksperymentów i kontrola warunków: ograniczenia danych (np. redukcja zbioru do 80% z powodu problemów pamięciowych PyTorch) wpływają na porównywalność i powinny zostać opisane jako istotne ograniczenie metodologiczne wraz z konsekwencjami. W części dot. YOLO Autor testuje przestarzałe modele detektora, których dziś nikt już nie używa.

Najbardziej interesująca część pracy (rozdział 5), dotycząca głębokich sieci neuronowych, jest bardzo lakoniczna i niespójna. Rozdział ten składa się głównie z ilustracji, a tekstu i porównania wydajności rozwiązań jest w nim bardzo mało (o ile porównaniem wydajności można nazwać pomiar czasu wnioskowania sieci na pojedynczym zdjęciu).

Podsumowując: Autor zebrał dane i pokazał, że potrafił uruchomić wiele narzędzi, ale samo ich porównanie jest metodologicznie niejednorodne, płytkich, chaotyczne i niezbyt jasne. Praca w obecnym kształcie bardziej pokazuje, że Autorowi udało się uruchomić ileś rozwiązań (w tym własne), niż rzetelnie porównać języki czy środowiska programowania sieci neuronowych jako takie.

3. Analiza literaturowa, dobór i sposób wykorzystania źródeł.

Bibliografia jest relatywnie obszerna (36 pozycji), ale zawiera głównie odniesienia do dokumentacji wykorzystanych narzędzi czy stron internetowych. Książki i artykuły (w tym materiałów z arXiv) występują sporadycznie. To nie jest samo w sobie krytycznym błędem (zwłaszcza w praktycznych pracach inżynierskich), natomiast proporcja źródeł stricte naukowych do „źródeł internetowych/dokumentacyjnych” jest nieodpowiednia.

Od strony edytorskiej bibliografia wymaga korekty formatowania, ponieważ w wielu pozycjach brakuje tytułów cytowanych źródeł, co utrudnia jednoznaczną identyfikację publikacji i weryfikację odwołań.

4. Trafność i spójność wniosków (krytyczna analiza osiągniętych wyników w odniesieniu do stanu wiedzy, możliwość dalszych kierunków badań).

We wnioskach Autor podkreśla m.in. znaczenie infrastruktury sprzętowej oraz fakt, że przy obliczeniach na GPU wpływ języka może być minimalny. Jest to słuszny wniosek i został częściowo poparty wynikami (szczególnie tam, gdzie dominuje czas uruchomienia gotowych modeli na GPU), ale uogólnienie jest w pracy sformułowane zbyt szeroko, bez dostatecznego rozdzielenia scenariuszy na własny kod (wtedy język, narzut środowiska, organizacja pamięci i wektoryzacja realnie mają znaczenie) oraz wykorzystanie gotowych, wysoko zoptymalizowanych bibliotek GPU (wtedy wąskim gardłem jest zwykle biblioteka/sprzęt, a nie język). Praca miesza te dwa aspekty, przez co końcowe wnioski są zbyt ogólne. Praca nie odpowiada rzetelnie na pytanie, które może być interesujące dla użytkownika: jaki język i jakie środowisko są najszybsze w implementacji współczesnych sieci neuronowych do przetwarzania obrazów.

Nierozstrzygnięte pozostają też kwestie poprawności interpretacji przyczyn różnic: np. stwierdzenia sugerujące wykorzystanie Tensor Core na podstawie „analizy kodu” są w praktyce

hipotezą, a nie wynikiem jednoznacznej weryfikacji pomiarowej. Tego typu tezy powinny być albo poparte twardą metodą (profilowanie, porównanie konfiguracji, kontrola środowiska), albo zapisane ostrożniej jako przypuszczenie.

Zalecane kierunki dalszych prac (np. sprawdzenie alternatywnych bibliotek i sprzętu, wpływ kwantyzacji, rozszerzenie o transformatory) są sensowne, ale część z nich powinna była być zastosowana w bieżącej pracy.

5. Układ i redakcja pracy (struktura formalna, przejrzystość, staranność edytorska, poprawność języka, wykorzystanie materiału ilustracyjnego).

Opiniowana praca liczy 50 stron (bez stron tytułowych i spisów rzeczy), jej układ jest prawidłowy, choć nie do końca spójny - układ rozdziałów nr 4 i 5, w których Autor opisuje przeprowadzone eksperymenty jest odmienny, a w gruncie rzeczy eksperymenty te mają podobny charakter. Język pracy jest na ogół poprawny, choć miejscami pojawiają się drobne potknięcia redakcyjne i stylistyczne, które jednak nie uniemożliwiają zrozumienia treści pracy. Dla lepszej ilustracji treści pracy Autor zamieścił w niej 36 rysunków i kilkanaście tabel. Do wielu z nich Autor nie odnosi się w tekście, część z nich nie ma także podanego źródła, a widać, że są skądś zaczerpnięte.

6. Ocena efektów uczenia się określonych dla pracy dyplomowej

Uważam, że pomimo wyżej wskazanych braków, efekty kształcenia przewidziane dla pracy inżynierskiej zostały osiągnięte.

7. Ocena zaangażowania i samodzielności pracy (samodzielność stawiania tez badawczych i rozwiązywania problemów, sumienność i terminowość wykonywania zadań, organizacja pracy, definiowanie kamieni milowych, raportowanie postępów)

Zaangażowanie Autora oceniam średnio, a terminowość nisko. Praca nad dyplomem trwała (z przerwami) około 20 miesięcy, a osiągnięte efekty są na granicy akceptowalności. Wprawdzie praca została wykonana przez Autora samodzielnie, ale wymagała znacznego i częstego wysiłku promotora w doprowadzeniu jej do aktualnego stanu.

☐

Inne uwagi

...

Opinia promotora w sprawie akceptacji pracy dyplomowej: pozytywna

Proponowana ocena pracy dyplomowej: trzy

.....

(data, podpis)