

mgr inż. Michał Wiraszka
(tytuł, imię i nazwisko recenzenta)

Recenzja 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 recenzji

1. Zgodność tytułu pracy dyplomowej z jej treścią

Tytuł pracy sugeruje porównanie wydajności języków programowania w kontekście sieci neuronowych. W rzeczywistości jednak praca w znacznie większym stopniu porównuje konkretne biblioteki i frameworki (PyTorch, TensorFlow, MATLAB, scikit-learn, Eigen, CUDA), różne poziomy abstrakcji implementacyjnej (implementacja od podstaw vs gotowe biblioteki), różne konfiguracje środowiskowe (CPU vs GPU, z różnym stopniem optymalizacji).

W wielu miejscach autor sam wskazuje, że porównywane implementacje różnią się stopniem dojrzałości, optymalizacji i źródłem (część kodu pochodzi z zewnętrznych repozytoriów). Powoduje to, że wnioski dotyczące „wydajności języków” są metodologicznie słabe lub wymagają daleko idących zastrzeżeń.

Ocena:

Tytuł jest częściowo zgodny z treścią, jednak upraszcza i nadinterpretuje faktyczny zakres badań. Bardziej adekwatny byłby tytuł akcentujący porównanie frameworków i środowisk obliczeniowych, a nie samych języków.

2. Wartość merytoryczna pracy

(identyfikacja problemu, sformułowanie celu, dobór i sposób wykorzystania narzędzi, rozwiązanie zadania badawczego)

Identyfikacja problemu i cel pracy

Problem porównania wydajności różnych technologii stosowanych w uczeniu maszynowym jest istotny i aktualny. Cel pracy jest jednak sformułowany zbyt ogólnie i nieprecyzyjnie:

nie jest jasno określone, co dokładnie oznacza „wydajność” (czas treningu, czas inferencji, zużycie zasobów, przepustowość?),

brak jednoznacznego rozróżnienia między wydajnością języka, biblioteki i sprzętu.

Dobór narzędzi

Dobór narzędzi jest szeroki i ambitny (Python, C++, Java, MATLAB, CUDA, różne sieci: regresja, MLP, CNN, YOLO). To niewątpliwa zaleta, ale jednocześnie źródło głównego problemu metodologicznego:

porównywane są niejednorodne implementacje (np. Java – implementacja własna, Python – dojrzałe biblioteki),

brak dowodu, że konfiguracje sieci, dane, hiperparametry i precyza obliczeń są identyczne,

w części eksperymentów (TensorFlow) środowisko działa na CPU zamiast GPU, co czyni porównanie niewspółmiernym.

Sposób wykorzystania narzędzi i rozwiązanie zadania

Eksperymenty zostały przeprowadzone poprawnie technicznie, jednak:

brakuje rygoru badawczego (brak powtórzeń, brak odchyleń standardowych, brak kontroli losowości),

pomiary czasu GPU nie są jednoznacznie opisane (brak informacji o synchronizacji),

część wyników (np. bardzo niska skuteczność sieci przy dużej liczbie epok) nie jest krytycznie przeanalizowana.

Ocena:

Wartość merytoryczna jest umiarkowana do dobrej, ale znaczco obniżona przez niedostateczną metodykę porównawczą i zbyt daleko idące interpretacje wyników.

3. Analiza literaturowa, dobór i sposób wykorzystania źródeł

Analiza literaturowa obejmuje:

dokumentację techniczną (NVIDIA, CUDA, cuDNN),

artykuły naukowe (arXiv),

materiały popularnonaukowe i internetowe (Wikipedia, Medium).

Z jednej strony pokazuje to praktyczne rozeznanie w narzędziach, z drugiej:

część źródeł ma niską wartość naukową i nie powinna stanowić podstawy definicji czy tez,

brak jest wyraźnego odniesienia wyników eksperymentów do literatury porównawczej (benchmarków, prac przeglądowych),

literatura jest wykorzystywana głównie opisowo, a nie analitycznie.

Ocena:

Dobór źródeł jest wystarczający, lecz ich wykorzystanie ma ograniczony charakter krytyczny i mogłyby zostać pogłębione.

4. Trafność i spójność wniosków

(krytyczna analiza wyników, odniesienie do stanu wiedzy, kierunki dalszych badań)

Wnioski są logicznie spójne z przedstawionymi wynikami, jednak:

są zbyt kategoryczne w stosunku do jakości danych eksperymentalnych,

często sugerują przewagę „języka”, podczas gdy faktycznie decydujące są biblioteki, backend GPU i konfiguracja,

brak rozróżnienia pomiędzy przypadkami poprawnej i niepoprawnej konfiguracji (np. TensorFlow na CPU).

Autor słusznie zauważa znaczenie sprzętu i akceleracji GPU, ale nie zawsze jasno oddziela:

obserwacje empiryczne,

hipotezy,

przypuszczenia (np. dotyczące użycia Tensor Core).

Kierunki dalszych badań są zarysowane jedynie pośrednio i mogłyby zostać sformułowane explicite.

Ocena:

Wnioski są częściowo trafne, lecz wymagają ostrożniejszego sformułowania i wyraźnych zastrzeżeń metodologicznych.

5. Układ i redakcja pracy

(struktura formalna, przejrzystość, staranność edytorska, język, ilustracje)

Praca ma:

poprawny i logiczny układ rozdziałów,

czytelne tabele i wykresy,

spójny styl narracji technicznej.

Jednocześnie występują:

drobne błędy merytoryczne w opisie sprzętu (np. typ pamięci GPU),
niejednoznaczne opisy metryk („efektywność” w YOLO),
momentami zbyt skrótowe opisy metod pomiarowych.

Język jest poprawny, techniczny i adekwatny do pracy inżynierskiej.

Ocena:

Układ i redakcja pracy są dobre, choć wymagają drobnych korekt precyzji i jednoznaczności opisów.

6. Ocena efektów uczenia się określonych dla pracy dyplomowej

Autor wykazuje:

umiejętność implementacji i uruchamiania zaawansowanych algorytmów ML/DL,
znajomość nowoczesnych narzędzi obliczeniowych,
zdolność do przeprowadzania eksperymentów i interpretacji wyników.

Słabiej natomiast widoczna jest:

umiejętność projektowania rzetelnego eksperymentu porównawczego,
krytyczna refleksja nad ograniczeniami zastosowanej metodyki.

Ocena:

Efekty uczenia się zostały osiągnięte w stopniu dobrym, choć nie w pełni na poziomie analitycznym oczekiwany od porównawczej pracy badawczej.

Inne uwagi

Praca ma charakter ambitny technicznie, ale zbyt szeroki zakres narzędzi utrudnił zachowanie rygoru porównawczego.

W przyszłości warto:

ograniczyć liczbę badanych technologii,
ujednolicić konfiguracje,
wprowadzić powtarzalność pomiarów i analizę statystyczną.

Praca spełnia wymagania formalne pracy inżynierskiej, ale jej wnioski należy traktować ostrożnie.

Przytoczone uwagi pozwoliły mi sformułować następujące pytania krytyczne dotyczące pracy:

I. Zakres pracy, cel i tytuł

1.

W tytule pracy mowa jest o „porównaniu wydajności języków”.

Proszę odnieść się do Rozdziału 1 (Wstęp, s. 7–8) i wyjaśnić:

jak autor definiuje „wydajność” w kontekście pracy,

dlaczego uznał, że badanie dotyczy języków, a nie przede wszystkim frameworków i bibliotek.

2.

Na s. 8 (Wstęp) autor deklaruje porównanie różnych języków w kontekście sieci neuronowych.

Czy z perspektywy wykonanych eksperymentów autor uważa, że wnioski rzeczywiście dotyczą języków, czy raczej konkretnych środowisk (PyTorch, TensorFlow, CUDA, MATLAB)?

Proszę to uzasadnić.

II. Metodyka badań i porównywalność eksperymentów

3.

W Rozdziale 4 (s. 36–42) oraz Rozdziale 5 (s. 43–55) porównywane są rozwiązania:

implementowane od podstaw (Java),

oparte o gotowe, silnie zoptymalizowane biblioteki (PyTorch, TensorFlow, scikit-learn).

Jak autor uzasadnia porównywalność takich rozwiązań i jaki wpływ ma to na interpretację wyników?

4.

W tabelach wyników (np. Tabela 2, s. 41) podane są pojedyncze czasy wykonania.

Proszę wyjaśnić:

ile razy wykonywany był każdy eksperiment,

czy badano powtarzalność wyników,

dlaczego w pracy nie podano odchyleń lub średnich z wielu pomiarów.

III. Regresja liniowa – Rozdział 3

5.

W Rozdziale 3 (s. 25–26) autor opisuje test regresji liniowej dla zbioru o rozmiarze 64 milionów elementów.

Proszę dokładnie wyjaśnić:

jakie operacje matematyczne były wykonywane,

czy była to jednorazowa operacja czy proces iteracyjny,

co dokładnie oznacza „czas obliczeń regresji” w Tabeli 1 (s. 26).

6.

Na s. 26 (Tabela 1) Python z NumPy osiąga znacznie dłuższy czas obliczeń niż inne języki.

Jak autor interpretuje ten wynik i czy sprawdzał, czy implementacja w NumPy była w pełni wektoryzowana?

IV. MLP i CNN – Rozdział 4

7.

W Rozdziale 4.3 (s. 38–41) porównywane są czasy uczenia MLP w różnych środowiskach.

Czy wszystkie modele miały:

identyczną architekturę,

tę samą liczbę epok,

te same dane wejściowe i batch size?

Jeśli nie, proszę wskazać różnice i wyjaśnić ich wpływ na wyniki.

8.

W Tabeli 2 (s. 41) pojawiają się różnice w dokładności modeli (np. 77% vs 94%).

Jak autor ocenia wpływ różnic jakości modelu na czas uczenia i czy był to czynnik brany pod uwagę?

9.

Na s. 41–42 autor sugeruje, że PyTorch mógł korzystać z Tensor Core.

Na jakiej podstawie technicznej sformułowano ten wniosek i czy był on weryfikowany (np. poprzez profilowanie lub ustawienia precyzji)?

V. CNN i TensorFlow – Rozdział 5

10.

W Rozdziale 5.2 (s. 47–48) autor stwierdza, że TensorFlow działał na CPU zamiast GPU.

Dlaczego mimo to wyniki TensorFlow zostały uwzględnione w porównaniach i jak wpływa to na końcowe wnioski?

11.

W Rozdziale 5.1–5.3 (s. 43–46) opisano eksperymenty na bardzo małym zbiorze danych (ok. 200 obrazów).

Jak autor uzasadnia:

dobór tak małego zbioru,

bardzo niską dokładność sieci (<20%),

oraz sensowność wniosków wydajnościowych w tej sytuacji?

VI. YOLO – Rozdział 5.4

12.

W Rozdziale 5.4 (s. 52–55) oraz w Tabelach 6–9 (s. 53–54) pojawia się pojęcie „efektywności”.

Proszę jednoznacznie zdefiniować:

co dokładnie oznacza ta metryka,

jak była liczona,

dla czego została wybrana zamiast standardowych miar (np. mAP).

13.

Na s. 52–53 autor wskazuje, że różne wersje YOLO nie są dostępne we wszystkich środowiskach.

Czy porównywane implementacje korzystały z:

tej samej wersji modelu,

tej samej rozdzielczości wejścia,

identycznego pre- i postprocessingu?

Jeśli nie – jak wpływa to na porównywalność czasów?

14.

W Rozdziale 5.4.1 (s. 55) autor pisze, że nie zawsze możliwy był pomiar czasów poszczególnych etapów.

Jakie są ograniczenia takiego podejścia i jak autor ocenia jego wpływ na wiarygodność wyników?

VII. Sprzęt i źródła

15.

W Rozdziale 2 (s. 22) opisano kartę graficzną RTX 4070 jako posiadającą pamięć „DDR5”.

Proszę odnieść się do tego opisu i wyjaśnić, czy jest on technicznie poprawny.

16.

W bibliografii (s. 56–57) znajdują się źródła takie jak Wikipedia i Medium.

Które fragmenty pracy opierają się na tych źródłach i czy – z perspektywy autora – mogłyby zostać zastąpione literaturą o wyższej randze naukowej?

Mając na uwadze powyższe proponuję ocenę dostateczną (3.0).

.....
(data, podpis)