dr hab. inż. Jakub Nalepa, prof. PŚ
Katedra Algorytmiki i Oprogramowania
Politechnika Śląska
Jakub.Nalepa@polsl.pl

Gliwice, 28 sierpnia 2023 r.

Recenzja rozprawy doktorskiej

Tytuł rozprawy: Data representations in generative modelling

Autor: Mgr inż. Kamil Deja

Promotor: Dr hab. inż. Tomasz Trzciński, prof. PW i UJ

1 Tematyka rozprawy

Recenzowana rozprawa doktorska Pana mgr. inż. Kamila Dei dotyczy zagadnień związanych z projektowaniem, analizą i wykorzystywaniem technik uczenia maszynowego w analizie i przetwarzaniu danych obrazowych, ze szczególnym uwzględnieniem metod modelowania generatywnego. Prace podjęte przez Doktoranta skupiały się zwłaszcza na procesie budowania reprezentacji danych, na analizie budowanych reprezentacji danych w przypadku zastosowania różnorodnych algorytmów generatywnych oraz na praktycznych zastosowaniach metod generatywnych, również w technikach uczenia ciągłego. Uważam, że tematyka rozprawy jest aktualna, zgodna z aktualnym nurtem badań i dotyczy istotnych aspektów uczenia maszynowego – zwłaszcza biorąc pod uwagę ogromną ilość danych (nie tylko obrazowych), które obecnie pozyskujemy. Opracowane algorytmy są generyczne i mogą znaleźć zastosowanie w rozwiązywaniu szeregu praktycznych problemów związanych z analizą danych, tworzeniem systemów wykorzystujących techniki uczenia ciągłego, wizją komputerową i analizą obrazów, np. medycznych czy satelitarnych.

2 Ocena strony merytorycznej rozprawy doktorskiej

Rozprawa doktorska jest napisana w języku angielskim i składa się z dziesięciu rozdziałów oraz bibliografii. **Rozdział pierwszy** (*Introduction*) rozpoczyna się od krótkiego wprowadzenia do tematyki rozprawy, w którym Autor uzasadnia istotność prowadzenia badań związanych z budowaniem reprezentacji danych oraz omawia praktyczne zastosowania metod generatywnych. Następnie, w podrozdziale 1.1. (*Research Questions*) Autor przedstawia trzy główne pytania badawcze, które będą rozważane w rozprawie – pytania badawcze są precyzyjnie sformułowane i nie są „trywialne". Bardzo dobrym pomysłem było podzielenie tych pytań, a w efekcie prac Doktoranta na trzy grupy, dotyczące *struktury*, *zastosowań* i *konsolidacji* (w kontekście uczenia ciągłego) wypracowywanych reprezentacji w modelach generatywnych (Rys. 1.1.1, str. 15 rozprawy). W kolejnym podrozdziale, *1.2. Thesis contribution*, Doktorant syntetycznie podsumowuje najistotniejsze aspekty pięciu prac wchodzących w skład cyklu publikacji będących osią rozprawy. Wszystkie z pięciu prac są publikacjami wieloautorskimi, a Autor precyzyjnie wskazuje swój wkład w powstanie tych prac – nie ma najmniejszych wątpliwości, że wkład Doktoranta w powstanie tych prac był kluczowy (ma to również odzwierciedlenie w pozycji Doktoranta na liście autorów – we wszystkich pracach jest pierwszym autorem), i że w pełni uzasadnione jest wykazanie tych prac w cyklu. Cztery artykuły, które stanowią trzon rozprawy, zostały opublikowane lub zaakceptowane do publikacji na prestiżowych konferencjach (NeurIPS 2022, ECML 2023, IJCNN 2021, IJCAI 2022 – wszystkie konferencje są notowane w rankingu CORE, odpowiednio: A*, A, B, A*; 200, 140, 70 i 200 pkt MEiN), z kolei jeden został opublikowany w czasopiśmie IEEE Access (100 pkt MEiN). Jest to również potwierdzenie tego, że metody i wyniki

zaprezentowane w powyższych pracach wpisują się w nurt najnowszych badań z zakresu technik generatywnych i budowania (i analizy) reprezentacji. W ostatnim podrozdziale rozdziału pierwszego, *1.3. Publications not included in the dissertation*, Autor przedstawia 11 prac, opublikowanych lub zaakceptowanych do publikacji na prestiżowych konferencjach (m.in. CVPR, Interspeech, ICONIP) i w czasopismach (m.in. Journal of Instrumentation), których był współautorem, a które nie zostały włączone w cykl publikacji będących trzonem rozprawy. Ponadto Doktorant podkreśla, że jest współautorem 130 publikacji, które powstały w ramach współpracy grupy ALICE Collaboration. Uważam, że dorobek publikacyjny Pana mgr. inż. Kamila Dei jest imponujący na tym etapie kariery naukowej.

Rozdział drugi jest (bardzo krótkim) wprowadzeniem teoretycznym do najważniejszych zagadnień związanych z tematyką rozprawy, zwłaszcza z metodami generatywnymi. Z kolei w **rozdziale trzecim**, Doktorant przedstawia syntetyczny przegląd literatury, który stanowi tło prac omawianych w rozprawie i pozwala lepiej zrozumieć najistotniejszy oryginalny wkład Doktoranta w rozwój metod generatywnych oraz technik budowania i analizy reprezentacji. W ostatnim podrozdziale rozdziału drugiego omówione zostały możliwości wykorzystania modeli generatywnych w fizyce wysokich energii, na przykład do szybkiej symulacji odpowiedzi kalorymetrów. Uważam, że jest to bardzo ciekawy przykład praktycznych zastosowań (których jest zdecydowanie więcej) algorytmów i technik, nad którymi przez ostatnie lata pracował Doktorant i jest to dowód na to, że wyniki tych badań są istotne i mogą znaleźć zastosowanie w innych dziedzinach nauki i przemysłu.

W **Rozdziałach 4–6**, Doktorant szczegółowo opisuje trzy z pięciu prac z omawianego cyklu publikacji [1, 2, 3]. Każdy z rozdziałów rozpoczyna się od krótkiego wprowadzenia, w którym Autor odnosi się do konkretnego zagadnienia (lub zagadnień), które były rozważane w omawianej pracy. W **rozdziale 4.** przedstawiony został model generatywny, w którym architektura autoenkodera (typu *Sinkhorn*) została rozszerzona o dodatkową sieć neuronową (*noise generator*), której celem jest zmapowanie szumu o zadanej charakterystyce na przestrzeń ukrytą wypracowaną przez autoenkoder dla danych rzeczywistych. Motywacja stworzenia takiego systemu została jasno omówiona, a sama architektura modelu generatywnego, która może zostać wytrenowana w podejściu *end-to-end*, jest „elegancka", przemyślana i szczegółowo uzasadniona w pracy. Opracowana metoda została eksperymentalnie zweryfikowana z wykorzystaniem zarówno klasycznych zbiorów benchmarkowych (MNIST, CelebA), jak również bardzo ciekawego zbioru symulowanych pomiarów kalorymetru. Autor przedstawia analizę ilościową i jakościową otrzymanych wyników oraz, w sposób przekonujący, wskazuje zalety opracowanej metody i porównuje ją z innymi technikami generatywnymi znanymi z literatury. **Rozdział 5** został poświęcony badaniom dotyczącym metod generatywnych wykorzystujących techniki dyfuzji – w ramach badań została zweryfikowana hipoteza mówiąca o tym, że takie modele mogą zostać podzielone na dwa główne komponenty – moduł generujący i odszumiający [3]. Badania te są istotne zwłaszcza w kontekście fundamentalnego zrozumienia tego, w jaki sposób działają metody generatywne oparte na technikach dyfuzji. Wyniki przekrojowych badań eksperymentalnych wykazały, że taki podział na część generującą i odszumiającą jest nie tylko możliwy i uzasadniony, ale może też pozwolić na poprawę jakości działania modelu generatywnego. Moim zdaniem, bardzo dobrym eksperymentem było wykazanie, że „siła" takiego systemu nie leży tylko w tym, że jest on oparty na dwóch modelach (w tym wypadku typu U-Net) – być może warto byłoby umieścić wyniki z podrozdziału 5.9.7. (z dodatku 5.9.) w głównym tekście rozprawy. W **rozdziale 6.**, Doktorant kontynuuje opis autorskich badań związanych z metodami wykorzystującymi techniki dyfuzji [3], tym razem w kontekście jednoczesnego budowania reprezentacji i modelu, np. klasyfikacyjnego. Szereg obserwacji eksperymentalnych, syntetycznie opisanych w tym rozdziale stanowi dobre tło, które pozwala zrozumieć to, w jaki sposób został zaprojektowany model łączący budowę reprezentacji danych i ich klasyfikację. Istotnym aspektem opracowanego systemu, w którym enkoder sieci typu U-Net jest współdzielony pomiędzy modułem generującym i klasyfikacyjnym jest to, że może być trenowany łącznie. Wszechstronne badania eksperymentalne pozwoliły na przekrojowe zweryfikowanie kilku istotnych aspektów opracowanej metody, które zostały szczegółowo omówione na str. 94 rozprawy. Bardzo ciekawym i pouczającym aspektem było zweryfikowanie opracowanych technik w innych scenariuszach eksperymentalnych (podrozdziały 6.6.4.–6.6.6.).

Rozdział 7 stanowi krótki (aczkolwiek treściwy) wstęp do uczenia ciągłego, w którym Autor omawia istniejące podejścia do tego problemu, zwłaszcza w kontekście tzw. katastroficznego zapominania i metod generatywnych w uczeniu ciągłym. Uczenia ciągłego dotyczą dwie kolejne prace zawarte w cyklu publikacji [4, 5], które są omawiane w **rozdziałach 8–9**. W **rodziale 8.**, Doktorant przedstawia autorski algorytm *BinPlay*, którego celem jest efektywne składowanie poprzednich obserwacji, które mogą być wykorzystane w trakcie treningu ciągłego. W pracy zaproponowano bardzo interesującą metodę „regenerowania" przykładów na podstawie wyznaczonego kodu binarnego, dzięki czemu – w przeciwieństwie do innych metod, które wykorzystują techniki powtarzania nauki dla poprzednio zgromadzonych przykładów treningowych przechowywanych w pamięci – wymagania pamięciowe metody *BinPlay* pozostają stałe (wraz ze wzrostem liczby zadań w scenariuszu uczenia ciągłego). Wyniki badań eksperymentalnych, przeprowadzonych dla trzech zbiorów benchmarkowych (MNIST, Fashion MNIST i CIFAR-10) jednoznacznie wykazały, że opracowana metoda pozwala na uzyskanie wyraźnie lepszych wyników zarówno od algorytmów buforujących poprzednie przykłady jak i tych opartych na wykorzystaniu modeli generatywnych. Warto podkreślić, że Doktorant przygotował materiał filmowy, który w przystępny i ciekawy sposób opisuje sposób działania algorytmu *BinPlay*. W **rozdziale 9.**, Autor szczegółowo omawia metodę *Multiband VAE*, której celem jest „konsolidacja" wiedzy wypracowanej w scenariuszu uczenia ciągłego, w której trening został podzielony na „lokalny" i „globalny". Warto zauważyć, że zaproponowano również nowy, bardziej realistyczny scenariusz uczenia ciągłego – jest to niezwykle istotne w kontekście przekrojowej walidacji metod uczenia ciągłego, które miałyby zostać wykorzystane w praktyce. Badania eksperymentalne objęły nie tylko klasyczne zbiory benchmarkowe (wykorzystanie wyłącznie zbiorów benchmarkowych wywołało mój pewien niedosyt w przypadku algorytmu *BinPlay*), ale także zbiór zawierający symulowane pomiary kalorymetru. Wyniki badań eksperymentalnych, które wykazały przewagę opracowanego podejścia nad innymi znanymi z literatury, zostały szczegółowo omówione przez Doktoranta.

W ostatnim **rozdziale 10.** Autor podsumowuje rozprawę. Doktorant precyzyjnie i jasno omawia aspekty badawcze związane z tematyką rozprawy, które mogą (i najpewniej będą) eksplorowane w najbliższym czasie.

Uważam, że prace, które zostały zawarte w cyklu publikacji będącym osią rozprawy zostały dobrane bardzo dobrze i odzwierciedlają przekrojowość badań prowadzonych przez Pana mgr. inż. Kamila Deję – wszystkie prace merytorycznie oceniam niezwykle wysoko. Na szczególną uwagę zasługuje też fakt, że udostępniono implementacje opracowywanych algorytmów, a wszystkie z omawianych scenariuszy eksperymentalnych były przemyślane, bardzo dobrze zaprojektowane i „rygorystyczne". Moim zdaniem zapewnienie reprodukowalności i przekrojowości eksperymentów jest obecnie niezwykle istotne, szczególnie w kontekście tzw. „kryzysu reprodukowalności" badań związanych z uczeniem maszynowym, z którym obecnie musimy się mierzyć [6, 7]. Nie mam wątpliwości, że Doktorant jest bardzo dobrze przygotowany do tego, żeby prowadzić dalsze badania związane (nie tylko) z uczeniem głębokim na światowym poziomie oraz żeby obiektywnie, rzetelnie i przekrojowo przedstawiać ich wyniki (w rozprawie wielokrotnie możemy przeczytać, że wybrane przykłady są „*non-cherry-picked*"). Rozprawa doktorska prezentuje zarówno szeroką wiedzę teoretyczną jak i praktyczną Doktoranta w dyscyplinie *Informatyka Techniczna i Telekomunikacja.*

2.1 Pytania i uwagi

W trakcie lektury rozprawy nasunęły mi się następujące pytania i uwagi:

1. Delikatnie niefortunnym sformułowaniem jest „*The PhD Candidate, as the first author, played a key role in developing and implementing the analysis and the proposed method.*" (w podrozdziale 1.2.2.). Być może lepiej byłoby napisać „*The PhD Candidate, as the first author, played a key role in developing and implementing the method, and in developing and conducting the analysis.*"?

2. Zauważyłem, że Autor ma tendencję do używania przymiotnika „*novel*" w odniesieniu do swoich prac. Zdecydowanie zgadzam się z tym, że wyniki badań i metody przedstawione w pracach

są nowatorskie, ale unikałbym używania takiego sformułowania (jeśli prace są „nowatorskie" to „bronią się" same – tak jak w przypadku prac Doktoranta, a jeśli nie są, to dodanie tego przymiotnika tego nie zmieni).

3. W pracy miejscami pojawiają się niejasne sformułowania – na przykład w Tabeli 3.1.1., czytamy (dla VAE), że dla tej metody możemy zaobserwować *„Limited performance on complex datasets"* (str. 33). Co dokładniej Doktorant ma na myśli przez *„limited"* i *„complex"* (i jak zdefiniować/ocenić, czy zbiór jest *„complex"*)?

4. Dlaczego Autor zdecydował się, w podrozdziale 3.1.5., na zaprezentowanie przykładowych wyników zawartych akurat w pracy Ding i in. (2018)? Warto byłoby także nieco dokładniej omówić zaprezentowany przykład (i to, co z niego wynika).

5. W rozprawie Doktorant umieścił rysunki, które czasami są trudne (np. Rys. 3.1.2.) lub praktycznie niemożliwe (np. Rys. 6.6.3.) do odczytania. Prezentacja wizualna nie tylko wyników eksperymentalnych, ale także opracowywanych metod jest bardzo istotna, bo może wyraźnie ułatwić (lub w tym wypadku utrudnić) czytelnikowi przyswojenie naistotniejszych treści, które chciałby przekazać autor tekstu. Warto podkreślić, że rysunki umieszczone w pracy przedstawiają bardzo istotne i ciekawe aspekty badań – tym większa szkoda, że niektóre z nich są „słabej jakości".

6. Pewnym mankamentem rozprawy jest to, że Autor nie umieścił w niej formalnych definicji metryk, które zostały wykorzystane w badaniach eksperymentalnych. Przykładowo, w przypadku metryki FID (Fréchet Inception Distance), Doktorant odwołuje się do pracy Heusela i in. (2017). Moim zdaniem lepiej byłoby formalnie zdefiniować wszystkie metryki użyte w rozprawie w oddzielnym podrozdziale, który podsumowywałby scenariusze eksperymentalne. Podobną uwagę można sformułować odnośnie zbiorów danych, które zostały użyte w pracy – umieszczenie ich opisu i charakterystyk (z podziałem na podzbiory treningowe, walidacyjne i testowe) w oddzielnym (pod)rozdziale nie tylko ułatwiłoby lekturę pracy, ale także mogłoby pomóc w uniknięciu powtarzania tych samych, lub bardzo podobnych opisów kilka razy.

7. Na str. 69 (w podrozdziale 5.7.2.) czytamy :*„In all cases, the performance of DAED is comparable to the DDGM if we replace the DAE with up to the 10% of the steps that correspond to log(SNR) is equal to around 4. For more complicated datasets like CIFAR10 and CelebA, fewer steps could be replaced. This effect could be explained by the fact that images in these datasets have three channels (RGB), and removing noise is more problematic."* – czy Autor podjął próby eksperymentalnej weryfikacji ten hipotezy (poza analizą wyników otrzymanych dla FashionMNIST, CIFAR10 i CelebA)?

8. Ocena jakościowa wyników działania (obrazów wyjściowych) jest nietrywialna nie tylko w przypadku metod generatywnych. Autor, na str. 70, zauważa: *„It seems that there is a sweet spot for perceptually appealing images that contain details and are "smooth" at the same time, see* $\beta_1 = 0.025$ *in Figure 5.7.3. However, as it is typically difficult to provide convincing arguments by staring at samples, we further propose to analyze quantitative measures."* – być może ciekawą opcją byłoby przeprowadzenie eksperymentu *Mean Opinion Score*, który mógłby pozwolić zobiektywizować analizę jakościową otrzymywanych obrazów (i mógłby być uzupełnieniem analizy ilościowej, jak np. w [8])?

9. Na str. 70 czytamy: *„In Appendix 5.9.7 we show that increasing the number of parameters of DDGMs to be comparable to DAED does not lead to* ***significant performance improvements.***" – czy Autor rozważał przeprowadzenie testów statystycznych, które mogłyby potwierdzić tę obserwację?

10. W podrozdziale 6.6., Autor podkreśla, że zweryfikowana zostanie odporność klasyfikatora (*„We measure the quality of a classifier to evaluate whether training together with a diffusion model improves the robustness of the classifier."*) – co dokładniej Autor rozumie przez *„robustness"* w tym kontekście?

11. Dlaczego w Tabelach 6.6.1. i 6.6.3. brakuje wyników eksperymentalnych dla niektórych metod z literatury dla wybranych zbiorów danych? Czy wynika to z tego, że wyniki eksperymentalne zostały bezpośrednio zaczerpnięte z odpowiadających im prac? Moim zdaniem taką informację warto byłoby zawrzeć w tekście rozprawy.

12. Na rys. 6.6.1. warto byłoby zaznaczyć, np. poprzez zaszarzenie tła panelu, obszar wykresu dla $\alpha \in [100, 500]$, ponieważ Autor wskazuje (na str. 96), że wartości obu metryk są wysokie dla tych wartości hiperparametru α, a więc te wartości hiperparametru α są istotne.

13. W obecnej wersji modelu (rozdział 6.5.), Autor wykorzystuje *average pooling* do łączenia cech ekstrahowanych przez różne poziomy sieci U-Net. Doktorant słusznie zauważył, w dodatku 6.8., że istnieją inne opcje, które mogłyby zostać wykorzystane do tego celu – czy oprócz podstawowych podejść, takich jak *min/max pooling*, Autor mógłby zaproponować inne techniki, które mogłyby zostać wykorzystane?

14. Na str. 89 Autor podkreśla, że – dla opracowanych reprezentacji – zostały wytrenowane modele klasyfikacyjne dla wszystkich 40 atrybutów ze zbioru CelebA. Wyniki klasyfikacji są jednak pokazane tylko dla 6 atrybutów (będących klasami w przypadku klasyfikacji) – w jaki sposób zostały wybrane te atrybuty? Podobne pytanie można sformułować dla przykładu zaprezentowanego na Rys. 6.8.3. – w jaki sposób zostały wybrane te atrybuty? Czy są z jakiegoś powodu „ciekawsze" od pozostałych?

15. Chciałbym poznać opinię Doktoranta na temat wykorzystania opracowanych metod w innych dziedzinach, np. w medycynie lub w analizie i przetwarzaniu danych satelitarnych. Czy możliwe byłoby wykorzystanie opracowanych metod uczenia ciągłego na urządzeniach brzegowych z ograniczeniami sprzętowymi (np. na pokładzie satelity obrazującego powierzchnię Ziemi, który miałby – w sposób ciągły – dostosowywać swoje działanie do nowych zadań)?

3 Ocena strony formalnej rozprawy doktorskiej

3.1 Ocena układu pracy i uwagi redakcyjne

Generalnie układ rozprawy doktorskiej jest poprawny i logiczny, a kolejne rozdziały wprowadzają czytelnika w najistotniejsze aspekty prac Doktoranta. Należy podkreślić, że napisanie rozprawy, która syntetycznie podsumowuje cykl publikacji, które są jej trzonem nie jest łatwe. Autor nie ustrzegł się pewnych uchybień, o których wspomniałem w innych częściach niniejszej recenzji. Oprócz tego, niektóre fragmenty rozprawy są kopią niektórych elementów publikacji z cyklu, co samo w sobie jest jak najbardziej uzasadnione, ale Doktorant skopiował te fragmenty z błędami, np. interpunkcyjnymi (np. w abstrakcie na str. 45 rozprawy czytamy: „*Our method outperforms competing approaches on a challenging dataset of simulation data from Zero Degree Calorimeters of ALICE experiment in LHC. as well as standard benchmarks, such as MNIST and CelebA.*". Czasami Autor odwołuje się do rysunków umieszczonych w innych pracach – np. na str. 63 czytamy „*Ho et al. (2020) notice that DDGMs can be beneficial for lossy compression, observing (Figure 5 in (Ho et al., 2020)) that most of the bits* (...)", a na str. 65: „*To be clear, we are not interested in how much each step of a DDGM contributes to the final objective (e.g., see Figure 2 in (Nichol and Dhariwal, 2021)) but rather* (...)". Moim zdaniem korzystne byłoby umieszczenie takich rysunków (lub rysunków inspirowanych wspomnianymi pracami) w rozprawie, aby ułatwić jej lekturę. Z kolei inne fragmenty rozprawy są nieco „nadmiarowe" i powtarzające się, np. te dotyczące istniejących algorytmów uczenia ciągłego czy opisu zbioru danych „*Real life CERN dataset*".

3.2 Ocena zastosowanego piśmiennictwa

Bardzo obszerna bibliografia zawiera około 280 pozycji, które zostały uporządkowane alfabetycznie. Dobór prac jest zdecydowanie poprawny i jest jednym z dowodów na to, że Autor rozprawy bardzo dobrze orientuje się w aktualnym nurcie badań związanych z budowaniem reprezentacji, metodami

generatywnymi oraz uczeniem głębokim w ogólności (w pracy możemy również znaleźć odniesienia do prac z 2023 r.). W trakcie analizy wykorzystanego piśmiennictwa zauważyłem jednak pewne uchybienia:

- W bibliografii możemy znaleźć wpisy, które się powtarzają (dlatego wyżej wspomniałem, że w spisie znajduje się „około" 280 pozycji), np. Goodfellow et al. (2014a) i Goodfellow et al. (2014b), Grathwohl et al. (2019a) i Grathwohl et al. (2019b), Kirkpatrick et al. (2017a) i Kirkpatrick et al. (2017b), Lecun et al. (1998) $\times 2$.
- Autor powinien zwrócić uwagę na poprawne wykorzystanie wielkich liter, np. Alice, gan, mcmc, Vae, Ieee – wystarczyłoby wykorzystać komendę `title={{My title}}`.
- W bibliografii są wpisy niekompletne, w których brakuje np. stron. Brakuje także spójności zapisu nazw konferencji – niektóre są zapisane tylko skrótem: IJCNN, inne pełną nazwą: IEEE/CVF Conference on Computer Vision and Pattern Recognition, a jeszcze inne pełną nazwą ze skrótem: 2021 International Joint Conference on Neural Networks (IJCNN).

3.3 Uwagi redakcyjne

Rozprawa jest napisana, w przeważającej części, starannie, a tekst rozprawy jest poprawny pod względem językowym i stylistycznym. W pracy można jednak zauważyć następujące uchybienia, które jednak nie wpływają negatywnie na odbiór tekstu – poniżej szczegółowo przedstawiam uchybienia i błędy, które zauważyłem w trakcie lektury rozprawy:

1. Autor naprzemiennie używa pisowni brytyjskiej i amerykańskiej, np. *regularisation* i *regularization* (str. 16), *parameterised* (str. 87) i *parameterized* (str. 61), „*e.g.*" i „*e.g.,*", „*behaviour*" i „*behavior*" i wiele innych. Najpewniej wynika to z tego, że niektóre z artykułów, które zostały zawarte w rozprawie były napisane z wykorzystaniem pisowni brytyjskiej, np. [1], a niektóre z wykorzystaniem pisowni amerykańskiej, np. [2].
2. Można zauważyć, że Autor w wielu miejscach niepoprawnie wykorzystuje komendy `cite` oraz `citep`. Przykładowo, zamiast „*As a result, to obtain discrete log-likelihoods, we consider the discretized (binned) Gaussian conditional likelihood Ho et al.(2020)*" na str. 29 powinno być „*As a result, to obtain discrete log-likelihoods, we consider the discretized (binned) Gaussian conditional likelihood (Ho et al.(2020))*".
3. W podpisie rysunku 4.4.2. możemy przeczytać „(...) *Our solution (bottom) generates* (...)" – w przypadku tego rysunku nie ma „*bottom*" (rysunek został zaczerpnięty z pracy [1], w której rzeczywiście wykorzystanie tego sformułowania jest uzasadnione, ponieważ elementy tego rysunku są w niej inaczej rozmieszczone).
4. Autor powinien zwrócić uwagę na spójność pisowni wielką i małą literą, m.in. w tytułach rozdziałów i podrozdziałów (np. *1.1. Research Questions* i *1.2. Thesis contribution*), ale także w tekście rozprawy (np. vae, VAE, wasserstein, Wasserstein, itp.) czy w podpisach rysunków i w samych rysunkach (np. t-SNE i TSNE na Rys. 3.1.3).
5. Doktorant nie ustrzegł się literówek i innych błędów językowych: „*singe*" (str. 26), „*of a standard GANs*" (str. 32), „*Our approaches alleviates*" (str. 35), „auoencoder" (podpis rysunku 3.1.3., str. 38), „(...) *of ALICE experiment in LHC. as well as* (...)" (str. 45), „*latente*" (str. 52), „*while it's energy and momenta* (...)" (str. 53), „*Well train DCGAN*" (str. 55), „*classfier*" (Rys. 6.4.2, str. 89), „*in the algorithm Algorithm 1*" (str. 93), „*we can focus on several training examples that can encoded* (...)" (str. 118), rozmiar obrazu dla zbioru CIFAR-10 powinien być 32×32 (str. 128), „*such split lead*" (str. 144), „*nerual*" (str. 157), „*leraning*" (str. 158).
6. W pracy występują niepotrzebne wcięcia, m.in. na str. 27 (po wzorze 2.4), na str. 51 (po wzorze 4.5).

7. Zamiast np. 5x5 lepiej byłoby napisać 5×5.

8. Zauważyłem błędy w niektórych skrótach, np. DEAD na str. 67 (zamiast DAED). Pomocne byłoby umieszczenie, np. we wstępie rozprawy, tabeli ze skrótami i symbolami wykorzystanymi w rozprawie – znacznie ułatwiłoby to lekturę pracy. Autor powinien też zwrócić uwagę na to, że każdy skrót powinien być zdefiniowany przy pierwszym użyciu, nawet jeśli jest „oczywisty" (np. MLP).

9. Pogrubienie najlepszych wyników we wszystkich tabelach, np. w tabelach 5.9.1 – 5.9.3 pozwoliłoby szybciej zidentyfikować czytelnikowi najlepsze algorytmy (lub najlepsze warianty algorytmów) porównywane w ramach tabeli.

10. W pracy możemy zauważyć niedomknięte nawiasy, np. w podpisie rysunku 5.9.1: „(c) CelebA)", brakujące lub niepoprawne znaki interpunkcyjne, np. „(...) explanations with a diffusion mode In this work, (...)" (podobnie na str. 137).

11. Doktorant powinien zwrócić uwagę na kolejność odnoszenia się do tabel i rysunków z tekstu – np. w sekcji *4.4. Experiments*, najpierw omawiany jest rysunek 4.4.1. (na str. 53), który pojawia się na str. 54, a dopiero później tabela 4.3.1 (na str. 55), która jest umieszczona na str. 53.

12. Na str. 75 (w podrozdziale 5.9.2.), Autor odsyła czytelnika do sekcji 3 – najprawdopodobniej jest to pozostałość z tekstu publikacji (w [2] istotnie odniesienie do sekcji 3 jest uzasadnione).

13. Wszystkie osie powinny być podpisane (np. na Rys. 9.6.2., 9.6.5., 9.6.6.).

Powyższe uchybienia nie wpływają na mój bardzo pozytywny odbiór rozprawy doktorskiej.

4 Konkluzja

Z pełnym przekonaniem stwierdzam, że recenzowana dysertacja Pana mgr. inż. Kamila Dei **spełnia** wymagania stawiane rozprawom doktorskim przez Ustawę – praca prezentuje ogólną wiedzę teoretyczną Dyplomanta w dyscyplinie Informatyka Techniczna i Telekomunikacja oraz umiejętność samodzielnego prowadzenia pracy naukowej, a przedmiotem rozprawy doktorskiej jest oryginalne rozwiązanie precyzyjnie zdefiniowanego problemu naukowego. W związku z powyższym, **wnioskuję o dopuszczenie mgr. inż. Kamila Dei do dalszych kroków procedury uzyskania stopnia doktora nauk technicznych**.

Ponadto, biorąc pod uwagę bardzo wysoką jakość prowadzonych badań, ich przekrojowość, innowacyjność i istotność opracowanych technik i algorytmów oraz bogaty dorobek publikacyjny Pana mgr. inż. Kamila Dei, **rekomenduję wyróżnienie rozprawy doktorskiej**.

Jakub Nalepa

Literatura

[1] K. Deja, J. Dubiński, P. Nowak, S. Wenzel, P. Spurek, T. Trzciński, End-to-End Sinkhorn Autoencoder With Noise Generator, IEEE Access 9 (2021) 7211–7219. doi:10.1109/ACCESS.2020.3048622.

[2] K. Deja, A. Kuzina, T. Trzciński, J. Tomczak, On Analyzing Generative and Denoising Capabilities of Diffusion-based Deep Generative Models, in: S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, A. Oh (Eds.), Advances in Neural Information Processing Systems, Vol. 35, Curran Associates, Inc., 2022, pp. 26218–26229.

[3] K. Deja, T. Trzciński, J. M. Tomczak, Learning data representations with joint diffusion models (2023). arXiv:2301.13622.

[4] K. Deja, P. Wawrzyński, D. Marczak, W. Masarczyk, T. Trzciński, BinPlay: A Binary Latent Autoencoder for Generative Replay Continual Learning, in: 2021 International Joint Conference on Neural Networks (IJCNN), 2021, pp. 1–8. doi:10.1109/IJCNN52387.2021.9534171.

[5] K. Deja, P. Wawrzyński, W. Masarczyk, D. Marczak, T. Trzciński, Multiband VAE: Latent Space Alignment for Knowledge Consolidation in Continual Learning, in: L. D. Raedt (Ed.), Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22, International Joint Conferences on Artificial Intelligence Organization, 2022, pp. 2902–2908, main Track. doi:10.24963/ijcai.2022/402.
URL https://doi.org/10.24963/ijcai.2022/402

[6] S. Kapoor, A. Narayanan, Leakage and the reproducibility crisis in machine-learning-based science, Patterns (Aug. 2023). doi:10.1016/j.patter.2023.100804.
URL https://www.cell.com/patterns/abstract/S2666-3899(23)00159-9

[7] S. Kapoor, E. Cantrell, K. Peng, T. H. Pham, C. A. Bail, O. E. Gundersen, J. M. Hofman, J. Hullman, M. A. Lones, M. M. Malik, P. Nanayakkara, R. A. Poldrack, I. D. Raji, M. Roberts, M. J. Salganik, M. Serra-Garcia, B. M. Stewart, G. Vandewiele, A. Narayanan, REFORMS: Reporting Standards for Machine Learning Based Science (2023). arXiv:2308.07832.

[8] B. Grabowski, M. Ziaja, M. Kawulok, P. Bosowski, N. Longépé, B. L. Saux, J. Nalepa, Squeezing nnU-Nets with Knowledge Distillation for On-Board Cloud Detection, CoRR abs/2306.09886 (2023). arXiv:2306.09886, doi:10.48550/arXiv.2306.09886.
URL https://doi.org/10.48550/arXiv.2306.09886