RADA NAUKOWA DYSCYPLINY
INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI WARSZAWSKIEJ

zaprasza na

OBRONĘ ROZPRAWY DOKTORSKIEJ

mgr. inż. Kamila Rafała Dei

która odbędzie się w dniu **9.10.2023 roku,** o godzinie **11.00** w trybie hybrydowym

Temat rozprawy:

„Data representations in generative modelling''

Promotor: dr hab. inż. Tomasz Trzciński, prof. uczelni – Politechnika Warszawska

Recenzenci: dr hab. inż. Jakub Nalepa, prof. uczelni– Politechnika Śląska

dr Eric Nalisnick – University of Amsterdam (Holandia)

dr hab. inż. Tomasz Kajdanowicz, prof. uczelni – Politechnika Wrocławska

Obrona odbędzie się w Audytorium Centralnym w gmachu Wydziału Elektroniki i Technik Informacyjnych (ul. Nowowiejska 15/19) oraz na platformie MS Teams pod linkiem: https://teams.microsoft.com/l/meetup-join/19%3ameeting_NjVlYzBkMmUtM2ZmNi00N2U1LWE5ZTUtMDc3NzljMDE5ZGI0%40thread.v2/0?context=%7b%22Tid%22%3a%223b50229c-cd78-4588-9bcf-97b7629e2f0f%22%2c%22Oid%22%3a%22cafbbddf-b922-42cc-bcda-6f80c1b14aeb%22%7d

Z rozprawą doktorską i recenzjami można zapoznać się w Czytelni Biblioteki Głównej Politechniki Warszawskiej, Warszawa, Plac Politechniki 1.

Streszczenie rozprawy doktorskiej i recenzje są zamieszczone na stronie internetowej: https://bip.pw.edu.pl/Postepowania-w-sprawie-nadania-stopnia-naukowego /Doktoraty/Wszczete-po-30-kwietnia-2019-r/Dyscyplina-informatyka-techniczna-i-telekomunikacja-dziedzina-nauk-inzynieryjno-technicznych/mgr-inz.-Kamil-Rafal-Deja

Przewodniczący Rady Naukowej Dyscypliny
Informatyka Techniczna i Telekomunikacja
Politechniki Warszawskiej
dr hab. inż. Jarosław Arabas, prof. uczelni

Data representations in generative modeling

This thesis presents a series of publications contributing to an in-depth understanding and development of generative models. We specifically focus on two recent approaches: generative autoencoders and diffusion-based generative models. We analyse how these methods build internal data representations and how they change when a model is retrained with additional data. We also propose several novel methods for generative modelling and their extensions to a continual learning setup.

In the first part of our work, we provide an overview of different generative autoencoders. We then introduce a novel model which allows for flexible encoding of examples into data representations, leveraging an additional neural network for sampling new data points.

In the second part, we move on to the recently proposed diffusion-based generative models. We begin by presenting an in-depth analysis of how the intermediate representations of images change with diffusion timesteps. Next, we introduce a novel joint model that demonstrates how the data representations generated by a diffusion model can be utilised to enhance performance in downstream tasks.

Finally, we extend our analysis to a continual learning scenario. Here, we show that generative approaches can be used as a universal method for continuous knowledge accumulation within models. To that end, we introduce two methods for continual generative modelling. In the first one, we propose a binary autoencoder that efficiently stores past experiences, while the second one is a method for the continuous alignment of data representations in the Variational Autoencoder's latent space.

Overall, our work contributes to the development of generative models through an in-depth analysis of their internal representations and novel ways of their application to real-life problems and continual-learning scenario.

Keywords: Generative Models, Continual Learning, Variational Autoencoder, Diffusion Models

Reprezentacje danych w modelowaniu generatywnym

W niniejszej pracy przedstawiamy serię publikacji poświęconą analizom i oryginalnym metodom modelowania generatywnego. W szczególności skupiamy się na sposobie budowania reprezentacji danych za pomocą powyższych metod. Badamy też jak wewnętrzne reprezentacje zmieniają się przy dotrenowywaniu modelu w oparciu o dodatkowe dane. W ramach przedstawionych prac proponujemy kilka nowych modeli generatywnych, wraz z ich rozszerzeniem do problemu uczenia ciągłego.

W pierwszej części pracy, dokonujemy przeglądu różnych autoenkoderów generatywnych. Następnie przedstawiamy nasz nowy model, który umożliwia elastyczne kodowanie przykładów do wewnętrznych reprezentacji, wykorzystując dodatkową sieć neuronową do próbkowania nowych obserwacji. W drugiej części, przechodzimy do zaproponowanych niedawno modeli generatywnych działających na zasadzie procesu dyfuzji. W pierwszej kolejności przedstawiamy analizy tego jak tymczasowe reprezentacje danych zmieniają się wraz z krokami dyfuzji. Następnie wprowadzamy nowy model łączny, za pomocą którego pokazujemy jak reprezentacje danych tworzone w procesie dyfuzji mogą być wykorzystane do poprawy wydajności w różnych zadaniach.

Następnie, rozszerzamy naszą analizę na problem uczenia ciągłego. W zagadnieniu tym pokazujemy że modele generatywne mogą być używane jako uniwersalna metoda do gromadzenia wiedzy napływającej w porcjach. W szczególności, przedstawiamy nasze dwie autorskie metody. W pierwszej wprowadzamy binarny autoenkoder, który wykorzystujemy do efektywnego przechowywania przeszłych doświadczeń. Natomiast w drugiej pracy, pokazujemy jak wykorzystać wariacyjny autoenkoder do ciągłej konsolidacji wiedzy poprzez uspójnianie ukrytych reprezentacji danych.

Podsumowując, w niniejszej pracy prezentujemy dogłębną analizę wewnętrznych reprezentacji modeli generatywnych oraz ich nowatorskie zastosowania w realnych aplikacjach, włączając w to problem uczenia ciągłego.

Sowa kluczowe: Modele Generatywne, Uczenia Ciągłe, Autoenkodery Wariacyjne, Modele Dyfuzyjne

dr hab inż. Tomasz Kajdanowicz, prof. uczelni
Katedra Sztucznej Inteligencji
Wydział Informatyki i Telekomunikacji
Politechnika Wrocławska

Wrocław, 4 września 2023

Recenzja rozprawy doktorskiej dla Rady Dyscypliny Naukowej Informatyki Technicznej i Telekomunikacji Politechniki Warszawskiej

Tytuł rozprawy: Data representations in generative modelling

Autor rozprawy: Kamil Deja

1 Analiza strony merytorycznej rozprawy

Rozprawę doktorską stanowi zbiór łącznie pięciu: czterech opublikowanych oraz jednego oczekującego na druk (zaakceptowanego), powiązanych tematycznie artykułów naukowych.[1]

1.1 Obszar problemowy

Praca doktorska skupia się na zadaniu uczenia reprezentacji danych w przestrzeniach posiadających strukturę użyteczną do wykorzystania w zadaniach docelowych (np. klasyfikacja, predykcja) z wykorzystaniem modeli generatywnych. Swą podstawę rozważań na ten temat wywodzi z możliwości stojących za architekturami głębokich sieci neuronowych takich jak klasyczny autokoder, wariacyjny autokoder, generatywnymi sieciami współzawodniczącymi (ang. Generative Adversarial Networks (GANs)) oraz dyfuzyjnymi modelami generatywnymi (ang. Diffusion-based Deep Generative Models (DDGMs)).

W rozprawie skupiono się na badaniu wewnętrznych mechanizmów różnych modeli generatywnych, starając się odpowiedzieć na trzy zasadnicze pytania badawcze. Pierwsze z nich dotyczy sposobu kodowania przykładów danych w reprezentacjach przydatnych dla modelowania generatywnego, drugie dotyczy możliwości wykorzystania reprezentacji danych nauczonej w sposób nienadzorowany przez modele generatywne poza zadaniem modelowania generatywnego, a ostatnie dotyczy wpływu retrenowania (dotrenowywania) modeli z wykorzystaniem dodatkowych danych na zmiany reprezentacji oraz możliwości łączenia wiedzy wprowadzanej do modelu generatywnego w oddzielnych zadaniach poprzez dopasowywanie reprezentacji (ang. alignment). W ogólności, przez dokładne zbadanie tych trzech pytań badawczych, rozprawa dąży do głębszego zrozumienia mechanizmów działania modeli generatywnych, ich ograniczeń oraz możliwości ich zastosowania w różnorodnych dziedzinach. Niewielką wadą tych pytań badawczych jest ich szeroki zakres i duże spektrum możliwości, co do rodzaju badań, które można na ich podstawie przeprowadzić. Chociaż każde z pytań jest niewątpliwie interesujące, nie dają one jednozacznego obrazu problemu, który planuje się rozwiązać. Na przykład,

[1] jedna z publikacji jest w druku w ramach konferencji ECML 2023, która odbywa się w terminie 18-22.09.2023

trudno jest określić, jakie kryteria powinna spełniać ”użyteczna reprezentacja” (np. takie jak inwariancje czy ekwiwariancje) lub co dokładnie oznacza ”użyteczność” w kontekście innych zadań niż generatywne. Mimo to, zastosowane podejście do sformułowania pytań badawczych w pracy jest zgodne z obecnym trendem, zwłaszcza gdy badania mają charakter empiryczny, a rezultaty powstają w ramach materialnych artefaktów naukowego procesu twórczego.

Pierwszy rozdział pracy stanowi kompleksowe wprowadzenie do głównych pytań i obszarów badawczych rozprawy, ukazując je w kontekście istniejących metod i potencjalnych zastosowań w uczeniu maszynowym i sztucznej inteligencji. Obejmuje on również omówienie pięciu opublikowanych wyników naukowych i wskazuje na dodatkowe publikacje, które nie zostały włączone do treści rozprawy.

Rozdział drugi wprowadza podstawowe koncepcje związane z poruszaną tematyką i przedstawia trzy różne rodzaje modeli generatywnych: (1) generatywne autokodery, jako szczególny przypadek autokoderów, posiadających pewne ograniczenia w kontekście modelowania generatywnego, i ich rozszerzenia do wariacyjnych autokoderów, (2) generatywne sieci współzawodniczące (GAN) z omówieniem ich zalet i wad, w tym braku zdolności uczenia reprezentacji dla konkretnych danych wejściowych, oraz (3) dyfuzyjne modele generatywne (DDGMs), które są postrzegane jako bardzo głębokie wariacyjne autoenkodery z określonym rodzajem posteriorów wariacyjnych. W rozdziale zawarte są wybrane detale techniczne związane z treningiem i modelowaniem tych sieci, a rozdział dostarcza analizę podstawowych podejść do generatywnego modelowania danych.

Rodział trzeci przedstawia pogłębioną analizę podejść związanych już z konkretnymi propozycjami przedstawionych w doktoracie rezultatów badawczych, w tym omówiono wybrane metody, które koncentrują się na reprezentacjach ukrytych z generatywnych autoenkoderów, generatywnych modeli dyfuzyjnych i ich zastosowań do różnych zadań, w tym zadań fizyki wielkich energii. Omówiono różne rozszerzenia i modyfikacje modeli generatywnych opartych na dyfuzji (DDGMs), które zaproponowane zostały w literaturze, w tym dla wielu modalności danych rozważanych łącznie. Podkreślono, że generatywne modele odgrywają kluczową rolę w bardzo różnych dziedzinach, od syntezy mowy po analizy biomedyczne. Ogólnie rzecz biorąc, rozdział zawiera dogłębną analizę i różnorodne zastosowania modeli generatywnych w różnych dziedzinach nauki.

Kolejne trzy rozdziały zawierają teksty opublikowanych, bądź zaakceptowanych do publikacji prac, które stanowią pierwszą część zbioru publikacji, t.j.,

rozdział czwarty - Deja, K., Dubiński, J., Nowak, P., Wenzel, S., Spurek, P., & Trzcinski, T. (2020). End-to-end sinkhorn autoencoder with noise generator. IEEE Access, 9, 7211-7219.

rozdział piąty - Deja, K., Kuzina, A., Trzcinski, T., & Tomczak, J. (2022). On analyzing generative and denoising capabilities of diffusion-based deep generative models. Advances in Neural Information Processing Systems, 35, 26218-26229.

rozdział szósty - Deja, K., Trzcinski, T.,& Tomczak, J. M. (2023). Learning Data Representations with Joint Diffusion Models. zaakceptowany na European Conference on Machine Learning.

Kolejny rozdział, siódmy, wprowadza nową perspektywę na rozważane dotąd problemy – uczenie reprezentacji w dynamicznych warunkach napływu danych (niestacjonarnośći danych). W przeciwieństwie do tradycyjnych modeli generatywnych, które są trenowane tylko raz na dostępnych danych, istnieje bowiem potrzeba ich dostosowywania się do nowych danych bez utraty wcześniej nabytej wiedzy. Poruszone jest zjawisko znane jako ”katastrofalne zapominanie”, gdzie model traci zdolność do wydajnego działania na starych danych, gdy jest przeuczony na nowych. W rozdziale omawia się taksonomię składającą się z (1) metod uczenia ciągłego (continual learning) wyodrębniając podejścia

bazujące na regularyzacji, sprowadzających się do penalizowania nauki i spowalniania procesu nadpisywania wcześniej nauczonej wiedzy, (2) metody bazujących na dynamicznych architekturach tworzących różne wersje modelu (głównie przyrostowe) dla różnych zadań, aby poprawić jego wydajność w scenariuszu wielozadaniowym oraz (3) metody bazujące na powtarzaniu zakładające, że zapominanie można pokonać jedynie przez stałe powtarzanie wcześniejszych przykładów danych. Rozdział prezentuje również w skrócie obecny stan wiedzy na temat metod uczenia ciągłego w modelowaniu generatywnym koncentrując się na zachowaniu zdolności do generowania z tej części rozkładu, która została wyestymowana ze starszych danych w strumieniu i wymieniając wszystkie główne prace stosujące różnorodne techniki, od regularyzacji wag po dynamiczne architektury i buforowanie danych.

Kolejne dwa rozdziały zawierają teksty opublikowanych prac, które stanowią drugą część zbioru publikacji, t.j.,

rozdział ósmy - Deja, K., Wawrzyński, P., Marczak, D., Masarczyk, W., & Trzciński, T. (2021). Binplay: A binary latent autoencoder for generative replay continual learning. In 2021 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

rozdział dziewiąty - Deja, K., Wawrzyński, P., Masarczyk, W., Marczak, D., & Trzciński, T. (2022). Multiband VAE: Latent Space Alignment for Knowledge Consolidation in Continual Learning. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22), (pp. 2902-2908).

Rozdział dziesiąty zawiera zwięzłe podsumowanie oraz dyskusję na temat przyszłości badań nad modelami generatywnymi i wpływu, jaki mogą one mieć na inne badania w dziedzinie uczenia maszynowego. Doktorant formułuje cztery bardzo interesujące pytania badawcze skupiające się na takich aspektach jak: czy model generatywny może pomóc w uczeniu reprezentacji w sposób użyteczny dla innych zadań uczenia ciągłego, czy reprezentacje danych wygenerowane przez generatywne modele dyfuzyjne mogą być korzystne w scenariuszach uczenia ciągłego, jakie zastosowanie mogą mieć reprezentacje uzyskane z generatywnych modeli dyfuzyjnych oraz w jaki sposób zaadoptować procesy dyfuzyjne by było możliwe uczenie reprezentacji na zasadzie autokodera. Rozdział zawiera interesujące propozycje dalszych badań umożliwiających poszukiwanie odpowiedzi na te pytania.

1.2 Ocena wyników oraz stopnia ich oryginalności

Oryginalne osiągnięcia badawcze przedstawione w doktoracie dotyczą zarówno opracowania nowych metod generatywnych umożliwiających otrzymywanie reprezentacji danych, jak i ich zastosowaniu w scenariuszu uczenia ciągłego.

Rozwiązania zaproponowane w pracy opierają się na innowacyjnych koncepcjach lub istotnych modyfikacjach w odniesieniu do metod znanych z literatury naukowej. Warto podkreślić dążenie do tworzenia rozwiązań, które charakteryzują się zarówno wysoką jakością, jak i efektywnością. W pracy można znaleźć wiele propozycji nowych rozwiązań, usprawnień, dodatków technicznych i implementacyjnych, które są wynikiem zgłębionej analizy i konsekwentej derywacji naukowej. Wśród uzyskanych rezultatów należy wymienić:

- wprowadzenie nowej architektury autokodera generatywnego nazwanego End-to-End Sinkhorn Autoencoder, w której proces kodowania próbek danych do ukrytych reprezentacji został oddzielony od procesu próbkowania nowych instancji, dodatkowa sieć neuronowa mapuje ustalony rozkład Gaussa na swobodnie kodowaną ukrytą reprezentację autokodera, a jako element regularyzacyjny użyta została aproksymacja Sinkhorna odległości Wassersteina, która wyrównuje zakodowane przykłady danych ze zmapowanym szumem,

- zaproponowanie hybrydowego, dwuczęściowego generatywnego modelu dyfuzyjnego, składającego się z odszumiacza i generatora, gdzie generuje się nowe cechy danych podobne do obserwowanych w danych treningowych (generator) i usuwa pozostały szum w sposób niezależny od danych (denoiser - deterministyczny autokoder),

- propozycję łącznego modelu dyfuzyjnego, w którym reprezentacje użyte są jako dane wejściowe dla klasyfikatora, a cały model jest optymalizowany łącznie na addytywnej funkcji kosztu (równanie 6.8) przy jednej parametryzacji do modelowania zarówno prawdopodobieństwa przykładu $p(x)$, jak i prawdopodobieństwa brzegowego $p(y|x)$,

- wprowadzenie binarnego autokodera jako metody efektywnego uczenia reprezentacji (BinPlay),inspirowane mechanizmem kompresji danych, w ramach uczenia ciągłego, gdzie modele generatywne służą jako źródło przykładów powtórzeniowych dla ciągle trenowanego klasyfikatora,

- wprowadzenie metody do ciągłego dopasowania wiedzy reprezentowanej w przestrzeni ukrytej autokodera wariacyjnego, gdzie aktualizacja ciągle trenowanego VAE jest podzielona na dwie części: (1) użycie nowego modelu do nauki osadzania nowych danych i (2) dopasowania (ang. alignment) wcześniejszych i nowych danych za pomocą dodatkowego projektora MLP działającego w ukrytej przestrzeni VAE.

Rezultaty te zostały opublikowane na bardzo dobrych koferencjach o renomie międzynarodowej (NeurIPS, IJCNN, IJCAI, ECML[2]) oraz w przyzwoitym czasopismie IEEE Access. Należy zaznaczyć, że publikacje wchodzące w skład rozprawy zostały napisane przez wielu autorów (kolejne publikacje miały 6, 4, 3, 5, 5 autorów). W tekście doktoratu została zawarta informacja na temat roli i udziału doktoranta w procesie powstania tych rezultatów, ale w żadnym z oryginalnych manuskryptów publikacji nie odnalazłem tych informacji. Doktorant, będący pierwszym autorem wszystkich prac, miał znaczący udział w planowaniu, konceptualizacji, implementacji, testowaniu i ocenie metod. Praca ta była realizowana na wartościowych stażach i we współpracy z bardzo dobrymi naukowcami. Całość potwierdza, że rola doktoranta jest wyraźna i stanowi główną siłę napędową powstania prac.

1.3 Zagadnienia dyskusyjne

Zawarte poniżej uwagi nie mają na celu krytykowania prezentowanej rozprawy, a stanowią zaproszenie do dyskusji i dalszej refleksji nad przedstawionymi w pracy zagadnieniami.

- *Pytania badawcze.* W rozdziale pierwszym pracy zostały zaprezentowane trzy pytania badawcze, na które odpowiedzi poszukuje przestawiony doktorat. Pojawiają się w nich określenia nieprecyzyjnie stawiające oczekiwania wobec metod uczenia reprezentacji. Proszę by Doktorant przeanalizował i odniósł się do tego, jakie kryteria powinna spełniać "użyteczna reprezentacja". Przykład dyskusji na ten temat jest zawarty w *Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. arXiv preprint arXiv:2104.13478.*

- *Dopasowanie (ang. alignment) za pomocą MLP*. W publikacji o Multiband VAE zaproponowana jest sieć tłumacząca, która uczy się wspólnego dopasowania oddzielnych przestrzeni ukrytych w zależności od identyfikatora zadania i od pojedynczej zmiennej ukrytej Z, w której wszystkie przykłady są reprezentowane niezależnie od swojego źródłowego zadania. Czy w związku z tym mechanizm translatora dokonuje projekcji jedynie przekształcającej przestrzeń czy raczej wyucza się dopasowania (przekształcenia) przestrzeni osadzeń jednego zadania względem drugiego? Jaka jest motywacja uczenia przekształcenia złożenia warstw liniowych z leaky relu?

[2] w druku

2 Analiza strony formalnej rozprawy

2.1 Ocena układu pracy i redakcji manuskryptu

Rozprawa doktorska napisana jest w języku angielskim i pod względem edytorskim jest opracowana starannie. Zawiera podziękowania, streszczenie w językach angielskim i polskim, spis treści, dwa rozdziały wprowadzające do pierwszej części zagadnień, jeden rozdział z przeglądem literatury, kolejne trzy rozdziały będące przedrukami artykułów, rozdział wprowadzający do tematyki drugiej części zagadnień, kolejne dwa rozdziały będące przedrukami artykułów oraz rozdział podsumowujący i prezentujący możliwe dalsze badania. Bibliografia zawiera 255 wymienionych i uporządkowanych alfabetycznie, wolnych od błędów pozycji bibliograficznych. Większość prac w bibliografii została opublikowana w ciągu ostatnich kilku lat, co świadczy o aktualności tematyki rozprawy na arenie międzynarodowej.

Układ rozprawy jest właściwy, aczkolwiek dostrzeżone niewielkie niedociągnięcia w rozprawie odnoszą się do organizacji rozdziałów pracy, w której rozdziały merytoryczne są przemieszane z wprowadzającymi tło (rozdział siódmy), a rozdział drugi i trzeci stanowią niepotrzebnie rozdzieloną zawartość, o czym mogą nawet świadczyć w zasadzie identyczne nazwy ich podrozdziałów. Są to jednakże nieistotne szczegóły, w znikomym stopniu wpływające na nawigację w pracy.

Użyta w pracy terminologia jest właściwa dla analizowanego obszaru badań, obejmującego uczenie maszynowe, modele generatywne, modele głębokie. Praca posiada 181 numerowanych stron.

2.2 Uwagi szczegółowe

Większość treści rozprawy jest poprawna pod względem językowym, stylistycznym i merytorycznym. W trakcie czytania zauważyłem jednak kilka drobnych błędów redakcyjnych i fragmentów, które w toku dodatkowej weryfikacji możnaby poprawić. Wiele razy pominięte zostały przecinki, w innych ich brakuje. Inne pomniejsze błędy:

- p.14 In this work, we focus on those inner-workings of different generative → In this work, we focus on the inner-workings of different generative
- p.14 continual-learning → continual learning
- p.16 noise → a noise
- p.22 forgetting of past data → forgetting past data

3 Konkluzja

Przedstawiona praca doktorska, autorstwa mgr. inż. Kamila Deji, nie tylko spełnia, ale wręcz przewyższa kryteria określone przez Ustawę z dnia 20 lipca 2018 r. - Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018 poz. 1668). Praca doktorska ukazuje głęboką znajomość dziedziny generatywnych technik uczenia maszynowego i pozwala wywnioskować o wysokich umiejętnościach samodzielnego prowadzenia pracy naukowej przez doktoranta. Praca koncentruje się na innowacyjnych odpowiedziach na kilka naukowych zagadnień, które łączy wspólny cel: identyfikacja efektywnych sposobów reprezentacji danych dla modeli generatywnych. Jest to praca mająca realny wpływ na dziedzinę, co można zaobserwować w licznych publikacjach autora w uznanych czasopismach oraz na międzynarodowych konferencjach z dziedziny informatyki. Przedstawione w pracy wyniki bardzo dobrze potwierdzają słuszność zgłoszonych przez autora koncepcji oraz efektywność ich zastosowania w praktyce. Wszelkie uwagi krytyczne, które pojawiły się w recenzji, nie wpływają na moją jednoznaczną i bardzo pozytywną ocenę tej dysertacji zarówno pod kątem oryginalności, jak i wartości merytorycznej. Określenie

zadań, proponowane innowacyjne metody ich rozwiązania, realizacja badań i kompetencje w formułowaniu wniosków dowodzą, że autor jest świetnie przygotowany do dalszej działalności naukowej. Bazując na tych przesłankach wnioskuję by mgr inż. Kamil Deja był dopuszczony do kolejnych etapów postępowania. Dodatkowo, ze względu na wysoką jakość publikacji i prestiżowych miejsc ich zamieszczenia, proponuję wyróżnienie jego pracy doktorskiej.

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) ×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`

Dr. Eric Nalisnick
Assistant Professor
Informatics Institute
University of Amsterdam

Address:
Science Park 900
1098 XH Amsterdam
email: e.t.nalisnick@uva.nl

Dissertation Review

Title: ***Data Representations in Generative Modeling***

Author of Dissertation: Kamil Deja, MSc.

1. Setting & Motivation

Generative models are an increasing important topic in modern AI and already have consequences for broader society. Models such as GPT and Stable Diffusion have demonstrated and forced us to consider the full range of benefits and dangers presented by generative models. For example, the benefits range from new creative technologies to automated assistants capable of a wide range of tasks to help with scientific discoveries. The harms are numerous as well: the generation of false news and mass propaganda, disrupting fair evaluation in education, and the careless violation of privacy and copyright. Autoencoder- and Diffusion-based architectures represent two of the most popular architectures in modern generative models. This dissertation does important work to understand (i) the information / structure encoded in the latent representations of these models, (ii) the wider utility of the representations, and (ii) the stability of the representation / its information when the model is fine-tuned on additional or different data. Topic (iii) will be especially important in the near future as people wish to apply foundation models to an increasingly wide range of tasks by way of finetuning. The dissertation respectively terms these three topics as “structure”, “applications”, and “consolidation”. Overall, I find the dissertation timely, well-motivated, and in conversation with the research frontier in terms of both applications and theory.

2. Novelty and Impact

2.1 End-to-End Sinkhorn Autoencoder

The contribution of the End-to-End Sinkhorn Autoencoder with Noise Generator (E2E-SAE) is to essentially define two encoders, one that maps random noise into the latent space and another that deterministically maps the sample into the same latent space. This presents a clever way to separate representation learning from the prior, as representation learning can happen on the deterministic path. Moreover, when combined with the stochastic encoder, the architecture still remains a proper generative model, as the deterministic encoder should generate outputs that follow

the aggregate prior implicitly defined by the other encoder. Moreover, this prior is also updated by the Sinkhorn loss for extra flexibility.

The model is clearly useful and competitive with the state-of-the-art, as demonstrated by the experiments that range from image to calorimeter generations. Thus, the model is clearly meets the bars for impact and novelty. My primary critique of the methodology is in regards to its ability to perform representation learning. I thought this was the primary motivation for the dual encoders, as described above, but I do not see any results that analyze the method's ability to perform good representation learning.

2.2 Analyzing Generative and Denoising Capabilities of Diffusion-Based Deep Generative Models

Diffusion models have achieved impressive success in their ability to generate high-quality, realistic, even seemingly creating images. While they are foundationally well-defined and understood as a continuous diffusion process, their computational pipeline consists of 1000+ steps. Unlike precursors models like VAEs and GANs, the subcomponents of diffusions models—if they even exist—-are hard to identify, let alone determine their role in image generation. This work aims to address this open problem, understanding the role of groups of steps of the diffusion.

In what I consider to be the highlight result of the dissertation, empirical analysis is conducted to show that diffusion models have two distinct phases: a generation phase and a denoising phase. The former is comprised of the first 80-90% of steps of the model, meaning that the model uses most of its capacity to translate noise into a structured image. The remaining 10-20% of steps are used to denoise the image in a fairly input-agnostic way. This perspective is further validated by successfully replacing the denoting steps with a denoising autoencoder, showing that this modular structure is indeed present. My only comment on this work is that, while the work on Generalized DAEs by Bengio et al. [2013] is mentioned, I thought their proposed 'walk-back' training procedure might be quite related to the proposed DAED model, as it also applies several steps of noise to the input before passing it into an autoencoder.

2.3. Learning Data Representations with Join Diffusion Models

This work continues the theme of probing and understanding diffusion models, further investigating the information contained in the intermediate states. Specifically, this work considers defining a joint distribution (so called 'hybrid model') that models both the features and a label, thereby doing both unsupervised density estimation as well as classification. The authors propose a clever, U-Net-based

model that sits in the middle of the diffusion process—not at the end, like previous approaches have done—and transports the intermediate states across the levels of the 'U'. The classifier can then use the features obtained at various points of the diffusion. Impressively, this model demonstrates all of the characteristics that a hybrid model promises: pure classification and generation, semi-supervised learning, and transfer learning. Moreover, the authors do this with a sensible, natural objective, not needing the 'hacks' that previous models have needed to balance generation and classification.

2.4. BinPlay: A Binary Latent Autoencoder for Generative Replay Continual Learning

Continual learning (CL)—the ability for a model to learn from a stream of new tasks, just as humans do, yet still retain past information—is a hallmark goal of artificial intelligence. Generative replay is a popular technique for CL, allowing the model to revisit previous important datapoints so as to not forget the information they provided. The work described in this section presents an interesting and practical take on the problem by formulating a generative model with a discrete latent space. Binary codes representing past data points can be stored and then fetched and passed to the decoder to obtain the full feature vector, which then can be passed to the classifier for replay. The experiments demonstrate strong classification performance in the CL setting and also report the memory footprint, verifying the motivation to use binary codes. While beyond the scope of the work, it would have been interesting to see the work better exploit the strong structure of binary codes to, e.g. organise the latent space into semantic concepts or some other hierarchy.

2.5. Multiband VAE

In this chapter, the dissertation describes a new procedure for unsupervised CL in which a local model is first trained on the task and then a global 'translator' model aggregates the local encodings. The motivation for such an approach is to cope with catastrophic forgetting—a key problem when performing CL. In short, the translator is trained with replay in order to prevent such forgetting. The experiments demonstrate that the model is competitive-to-superior to other SOTA generative CL models (in terms of FID score). This demonstrates the utility of this sensible approach. My only critique is that I wonder if such an approach will be less sample efficient, especially for later tasks, as the model cannot use information from previous tasks to help with learning later tasks.

3. Evaluating the Written Document

Overall, the document is in the appropriate format, providing the necessary context and background information before diving into specific works in the middle chapters. The document closes with a summary and research questions for future work, which are appropriate, if a bit short-sighted. I have only two remarks for improvements: Regarding the E2E-SAE model, I found the description rather algorithmic in that it describes what the model it doing in a procedural manner. However, I think a more conceptually 'cleaner' presentation is to separately define the generative model, the inference model, and the training procedure. This makes clear when one could make alternative choices for each of these building blocks. Secondly, the switch to continual learning is somewhat abrupt. Perhaps this could be better integrated into the background on generative models. Alternatively, it could be kept in its current place but with more, earlier discussion about the connections between generative modelling and continual learning (what's currently in 7.2). Yet, in summary, I found the dissertation a pleasure to read. See the appendix for detailed comments on minor points.

4. Conclusions

The reviewed dissertation of Kamil Deja, MSc., meets the requirements for doctoral dissertations by the Act on Scientific Degrees and Academic Title of 14 March 2003 (Journal of Laws No. 65) as it presents novel, impactful concepts that push forward the discipline of Computer Science. This fact is further supported by the prestigious and selective venues in which the dissertation's material has already been published (e.g. *Neural Information Processing Systems* 2022, International Joint Conference on Artificial Intelligence 2022). Any critical remarks presented above should not detract from my incontrovertibly positive assessment. **I request that the doctoral degree be awarded to Kamil Deja, MSc, and due to the prestige and impact the work has already gathered, I recommend the degree be awarded with honors.**

Please do not hesitate to contact me at *e.t.nalisnick@uva.nl* if you have any further questions.

Sincerely,

Eric Thomas Nalisnick, PhD

Appendix: Detailed Remarks

p 13: *"Normalising flows (Rezende and Mohamed, 2015) and Glows (Kingma and Dhariwal, 2018) are explicitly trained to map original data samples into a lower-dimensional manifold through invertible operations"*: Firstly, Glow is a type of normalizing flow, and secondly, I find the statement "map…samples into a lower-dimensional manifold" a bit misleading since flows preserve dimensionality. This does happen by introducing inductive biases such as multi-scale architectures, but the core mathematics does not allow this.

Eq 2.1, 2.2, 2.3, 2.4, and others: loss functions / optimization objectives should usually be written as functions of the parameters that one would optimize (theta and phi, in this case 2.1)

p 26 *"we cannot simply propagate a gradient through a random node."*: This is an overstatement since the reparameterization trick wasn't the first method to differentiate through samples. For instance, Kingma & Welling could have used the existing score function estimator (REINFORCE). The contribution is more of providing a practical, low-variance gradient estimator (with the plus that, when combined with the variational approximation, made the overall computation graph look like an autoencoder).

p 28, 48, 61, 85, and others: You give several references in NAME [YEAR] format but look like they should be in [NAME, YEAR] format.

p 34: The reference to my paper should be ICLR 2019, not 2018. Moreover, I'm not sure my paper is the best one to cite when describing the effect of the aggregated posterior. My contribution is more about the aggregated posterior's effect on OOD detection. I first learned about the aggregated posterior from the VAE papers that re-write the ELBO to include a KL term between the aggregated and per-data-point posteriors, for the purposes of representation learning. This was also discussed in the VampPrior paper, which pre-dates my ICLR 2019 paper.

Eq 4.3 and 4.4: Could be written in one line instead of two.

Eq 9.1, 9.2, 9.3: These expressions should be set equal to something