Autouzupełnianie chmur punktów 3D wspomagane sztuczną inteligencją

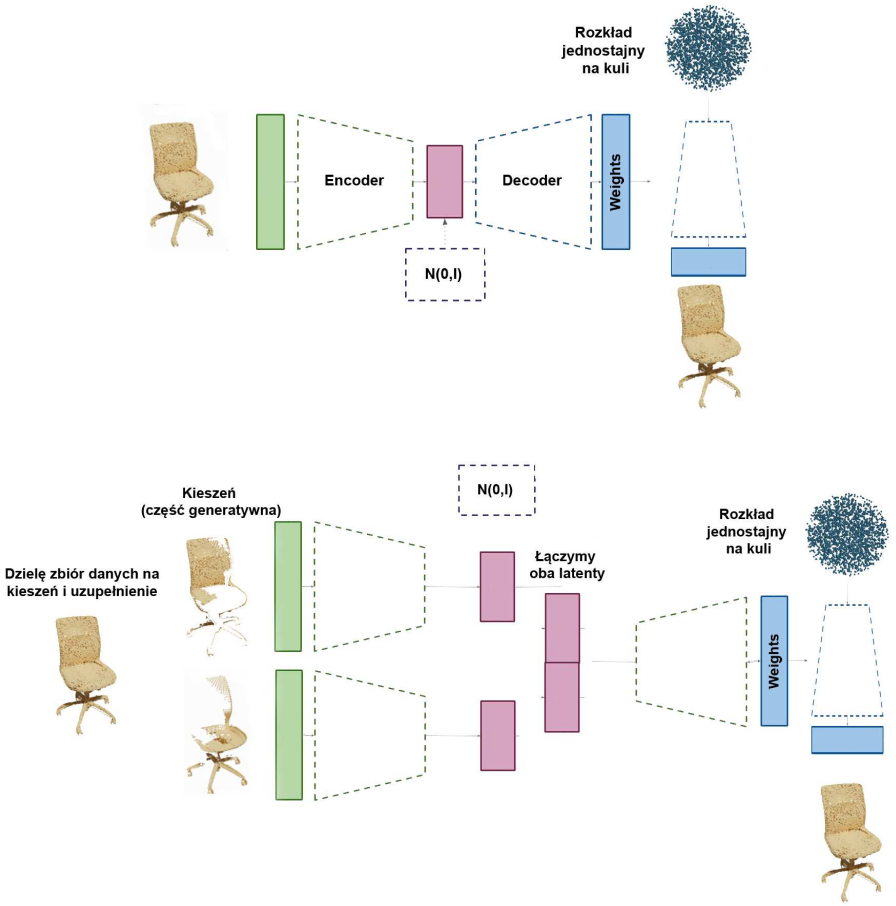
1 Cel projektu

Celem naszego projektu będzie zbudowanie modelu, stanowiącego część systemu autouzupełniania chmur punktów 3D (*3D Point Clouds*). Motywacją dla realizacji takiego zadania jest działanie systemów typu LiDAR [1], które udostępniają do „zczytywania" jedynie cząstkowe dane, które należy w odpowiedni sposób uzupełnić. Jako punkt wyjścia zamierzamy wykorzystać algorytm *HyperCloud* z artykułu [2], który wykorzystuje koncepcję tzw. hipersieci (*HyperNetwork*), czyli większej architektury tworzącej mniejszą sieć neuronową dedykowaną do realizacji konkretnego zadania – w tym przypadku wynikowa sieć ma generować chmury punktów reprezentujące obiekty 3-wymiarowe, podobne do elementów ze zbioru uczącego. Schemat działania takiego modelu przedstawiony jest na górnej części rys. 1 – na wejściu zadajemy chmurę punktów, która po przejściu przez architekturę typu autoenkoderowego tworzy wagi wyjściowej sieci, umożliwiającej (po nauczeniu) generowanie docelowych obiektów z 3-wymiarowego szumu.

Naszym celem jest adaptacja powyższego podejścia do sytuacji, gdy mamy zadany jedynie fragment obiektu 3D (np. krzesła), zaś pozostała jego część pozostaje „niewidoczna". Docelowo chcemy otrzymać model, który potrafi uzupełniać te braki oraz posiada charakter generatywny, tzn. umożliwia losowe generowanie fragmentów „pasujących" do niewidocznej części zadanego obiektu. Mówiąc bardziej obrazowo, nasze rozwiązane będzie zawierać pewnego rodzaju „kieszeń" (stąd nazwa *HyperPocket*), którą należy w odpowiedni (ale losowy) sposób wypełnić.

Należy podkreślić, iż prezentowany projekt posiada charakter długofalowy. Kolejnym rozważanym etapem (nie wchodzącym jednak w zakres niniejszego wniosku) będzie adaptacja uzyskanego rozwiązania do przypadku scen wewnętrznych, składających się z (potencjalnie) wielu pojedynczych obiektów. Naturalnym podejściem do takiego zagadnienia jest wycinanie poszczególnych kawałków sceny i traktowanie ich jako dane wejściowe dla zaimplementowanego wcześniej modelu. Oczywiście, efekt końcowy będzie wówczas zależał od odpowiedniego doboru zbioru uczącego (należy mieć na uwadze, że wycięty fragment, który ma reprezentować pojedynczy obiekt, może również zawierać kawałki innych obiektów).

Wyrażamy nadzieję, iż w wyniku naszej współpracy (którą chcielibyśmy w przyszłości rozszerzyć o ośrodek zagraniczny oraz ośrodek z sektora przemysłowego) powstanie uniwersalny system przetwarzania chmur punktów, umożliwiający tworzenie tzw. *cyfrowych bliźniaków* (zob. sekcja 3).



Rysunek 1: Górna część rysunku pokazuje model używany w pracy [2]. Dolna część rysunku reprezentuje schemat nowej architektury realizującej proces autouzupełniania.

2 Implementacja

Implementacja projektu zostanie przeprowadzona na repozytorium *ShapeNet*[1] [4]. Docelowy model składać się będzie z następujących części:

[1] Na etapie adaptacji modelu do przypadku scen wewnętrznych (niewchodzącym w zakres niniejszego projektu) zamierzamy korzystać z repozytorium *ScanNet* [3].

- dwóch enkoderów, z których pierwszy będzie prowadzić w przestrzeń ukrytą (*latent*) z zadanym apriori rozkładem $N(0, I)$,
- dekodera określonego na produkcie kartezjańskim przestrzeni ukrytych obu enkoderów, zwracającego parametry docelowej sieci neuronowej, mającej za zadanie generować docelowe obiekty z 3-wymiarowego szumu (rozkładu jednostajnego na kuli).

Każdy obiekt będziemy dzielić (potencjalnie wielokrotnie) na dwie części, z których jedna będzie stanowić część niewidoczną (czyli „kieszeń"), traktując je jako dane wejściowe dla poszczególnych enkoderów. W procesie uczenia (zob. dolna część rys. 1) będziemy stosować procedurę analogiczną do tej zaprezentowanej w pracy [2], z tą różnicą, że w obecnym modelu będziemy uwzględniać dodatkowy enkoder, uzupełniający docelową przestrzeń ukrytą o część niegeneratywną.

3 Uczenie maszynowe a cyfrowy bliźniak

Naszym długofalowym celem jest nawiązanie współpracy z ośrodkiem zagranicznym oraz ośrodkiem z sektora przemysłowego, które interesują się badaniami nad koncepcją *cyfrowego bliźniaka* [6]. Niniejszy projekt jest pierwszym krokiem w budowaniu takiej współpracy.

3.1 Koncepcja *Przemysłu 4.0*

W ciągu ostatniej dekady koncepcja tzw. *Przemysłu 4.0* [5] przekształciła się od nieśmiele nakreślanej idei do pełnego programu badawczo-rozwojowego. Skupia się on na cyfryzacji oferowanych przez przedsiębiorców towarów i usług oraz tworzeniu nowych, opartych o innowacje modeli biznesowych [5]. Jednymi z najważniejszych komponentów platformy *Przemysłu 4.0* są:

- zapewnienie niezawodnej oraz szybkiej komunikacji pomiędzy parkiem maszynowym, a ich użytkownikami;
- wsparcie procesów projektowania, produkcji oraz zarządzania w całym cyklu życia danego produktu lub usługi;
- wykorzystanie tzw. *cyfrowych bliźniaków* [6];
- wykorzystanie zaawansowanych technologii asystywnych [7].

Istotnym jest tutaj fakt, że tak ambitny projekt wymaga olbrzymiej interdyscyplinarności i współpracy ekspertów z wielu dziedzin. Ze względu na innowacyjność całej koncepcji należy tutaj wymienić kluczową rolę jaką ma do spełniania świat nauki. Oszacowanie możliwości stworzenia danego rozwiązania, jego zaprojektowanie, zbadanie a następnie przetestowanie wymaga współpracy naukowców reprezentujących różne, nieraz odległe dziedziny wiedzy oraz ekspertów z różnych gałęzi inżynierii, matematyki, informatyki czy specjalistów od interakcji człowiek–maszyna.

3.2 Koncepcja *cyfrowego bliźniaka*

W celu urzeczywistnienia koncepcji *Przemysłu 4.0* [5], jednym z najistotniejszych jej elementów jest tzw. *cyfrowy bliźniak* [6]. Cyfrowy bliźniak jest cyfrowym odwzorowaniem realnie istniejącego urządzenia lub obiektu. Takie odwzorowanie może skupiać się na cechach fizyko-chemicznych danego obiektu, takich jak jego skład chemiczny, lub też na jego zewnętrznych charakterystykach, takich jak jego wysokość, szerokość czy kolor. W zależności od tego o jakim obiekcie mówimy i jak dokładnym jego odwzorowaniem dysponujemy, potencjalny zakres jego zastosowania takiego modelu może się znacząco poszerzyć. W tym kontekście możemy również dysponować modelem całej sceny 3D, np. zamkniętego pokoju wraz ze wszystkimi sprzętami lub hali produkcyjnej wraz z poszczególnymi robotami i innymi elementami wyposażenia [8]. Potencjał wykorzystania cyfrowych bliźniaków jest olbrzymi – mogą być one używane np. do dokonywania inspekcji inżynierskich [9] lub śledzenia etapów inwestycji budowlanej.

3.3 Cyfrowy bliźniak a chmura punktów 3D

Największymi obecnie przeszkodami w tworzeniu *cyfrowych bliźniaków* jest czas ich produkcji, koszty z nią związane oraz ich finalna jakość [9].

Istnieje kilka technik, które można wykorzystać do tworzenia fotorealistycznych modeli istniejących obiektów rozumianych jako odwzorowanie kształtu i kolorystyki danego obiektu lub ich kolekcji w postaci sceny. W tym drugim kontekście bardzo istotne jest dokładne oddanie relacji przestrzennych obiektów znajdujących się w pewnej odległości i rotacji względem siebie w przestrzeni 3D. Jednym z szeroko wykorzystywanych sposobów uzyskiwania takiej informacji przestrzennej jest skanowanie laserowe, w tym wykorzystanie systemu LiDAR lub innych pokrewnych technologii [1]. Pozwala ono na uzyskanie modelu obiektu w postaci m.in. chmury punktów 3D. Niestety, analiza takiej chmury, ze względu na jej rozmiar (ilość wygenerowanych punktów), częstą monochromatyczność oraz niekompletność zebranych danych stanowi otwarty problem badawczy.

Dwa główne kierunki badań, których rozwinięcie może stanowić remedium na ten problemu, stanowią:

- automatyczna segmentacja trójwymiarowej sceny poprzez detekcję osadzonych w niej obiektów 3D [11, 12];
- możliwość automatycznego „dosztukowywania” brakujących fragmentów danej sceny 3D [10].

Biorąc pod uwagę postęp poczyniony w tej dziedzinie w ciągu ostatnich lat, wydaje się, że najpotężniejszym narzędziem, jakie można tutaj zastosować, są techniki stosowane w uczeniu maszynowym [11, 12], co stanowi wystarczające uzasadnienie i motywację dla podjęcia prezentowanego w niniejszym wniosku problemu badawczego.

Literatura

[1] Minos-Stensrud et al., *Towards Automated 3D reconstruction in SME factories and Digital Twin Model generation*, ICCAS 2018.

[2] Spurek et al., *Hypernetwork approach to generating point clouds*, ICML 2020.

[3] Dai et al., *ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes*, CVPR 2017.

[4] Wu et al., *3D ShapeNets: A Deep Representation for Volumetric Shape Modeling*, CVPR 2015.

[5] *Polska Platforma Przyszłości,* `https://przemyslprzyszlosci.gov.pl/`

[6] Saddik, *Digital Twins: The Convergence of Multimedia Technologies*, IEEE MultiMedia 25 (2018), 87–92.

[7] Möncks et al., *Worker assistance systems: Understanding the human perspective*, Institute of Manufacturing, University of Cambridge, UK, 2020.

[8] Sommer et al., *Automatic Generation of Digital Twin Based on Scanning and Object Recognition*, Advances in Transdisciplinary Engineering, 645–654.

[9] Tadeja et al., *PhotoTwinVR: An Immersive System for Manipulation, Inspection and Dimension Measurements of the 3D Photogrammetric Models of Real-Life Structures in Virtual Reality*, 2019, arXiv.

[10] Litany et al., *ASIST: Automatic semantically invariant scene transformation*, Computer Vision and Image Understanding 157 (2017), 284–299.

[11] Qi et al., *ImVoteNet: Boosting 3D Object Detection in Point Clouds With Image Votes*, CVPR 2020.

[12] Qi et al., *Deep Hough Voting for 3D Object Detection in Point Clouds*, ICCV 2019.