

Field of study: **Artificial Intelligence**  
Speciality: ---

## MASTER THESIS

### **A Method for Image Generation Using Conditional Multi-Views**

Eryk Wójcik

Supervisor  
**dr hab. inż. Maciej Zięba**

Machine Learning, Generative Models, Diffusion



## Streszczenie

TESTTTTTT Dodaj streszczenie pracy w języku polskim. Staraj się uwzględnić wymienione na stronie tytułowej słowa kluczowe. Uwaga przedstawiony rekomendowany szablon dotyczy pracy dyplomowej pisanej w języku angielskim. W przeciwnym wypadku, student powinien samodzielnie zmienić nazwy „Chapter” na „Rozdział” itp stosując odpowiednie pakiety systemu L<sup>A</sup>T<sub>E</sub>X oraz ustawienia w pliku *latex-settings.tex*.

## Abstract

Streszczenie w języku angielskim.



# Spis treści

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Introduction</b>                                | <b>1</b> |
| 1.1      | Problem Statement . . . . .                        | 1        |
| 1.2      | Thesis Objectives . . . . .                        | 1        |
| 1.3      | Thesis Outline . . . . .                           | 1        |
| <b>2</b> | <b>Related Work</b>                                | <b>3</b> |
| 2.1      | SLAM-based solutions . . . . .                     | 3        |
| 2.2      | Image-prompt based generative models . . . . .     | 3        |
| 2.3      | Adapters for multi-view image generation . . . . . | 3        |
| <b>3</b> | <b>Politechnika Wrocławska</b>                     | <b>5</b> |
| 3.1      | Tytuł podrozdziału . . . . .                       | 5        |
| 3.1.1    | Tytuł podpodrozdziału . . . . .                    | 6        |
| 3.2      | Tytuł podrozdziału 2 . . . . .                     | 6        |
| <b>4</b> | <b>Conclusions and Future</b>                      | <b>9</b> |



# 1. Introduction

W pracy formułuje się cele o charakterze badawczym wymagające doboru i zastosowania metod badawczych, wykorzystując wiedzę teoretyczną oraz naukową. Wskazane jest przedstawienie, co nowego jest zaproponowane w pracy oraz podanie ograniczeń i słabych/mocnych stron opracowanego rozwiązania (jeżeli dotyczy). Rozdział wprowadzający powinien służyć czytelnikowi do zrozumienia celu pracy.

## 1.1. Problem Statement

W tej sekcji student powinien przedstawić bliżej problem, którym chce się zmierzyć. Jasno zdefiniuj problem badawczy. Podaj swoje cele, zadania i pytania badawcze. Wyjaśnij znaczenie badania. Określ ograniczenia badań.

## 1.2. Thesis Objectives

W tej sekcji powinny zostać przedstawione konkretne działania, które określają pracę studenta w celu rozwiązania problemu.

## 1.3. Thesis Outline

Zarysuj strukturę swojej pracy dyplomowej. Ogólnie przedstawienie pracy. Przykładowo: „Praca dzieli się na 7 rozdziałów (...)”. Rozdział [3](#) dotyczy (...). Temat został rozwinięty w [3.1](#).





## 2. Related Work

In this chapter, we review existing approaches to multi-view image generation and novel view synthesis. We first discuss traditional SLAM-based methods of novel view synthesis (Section X.1), followed by image-prompt based generative models for view synthesis (Section X.2), with particular emphasis on adapter based methods to fine-tune diffusion models to generate views from a single image (Section X.3). Finally, we analyze the limitations of current methods and identify research gaps that our work aims to address (Section X.4).

The general idea of novel view synthesis is to generate a new view of a scene from a set of multi-view images. The main challenge is to handle the scene's 3D structure and the camera's motion.

### 2.1. SLAM-based solutions

### 2.2. Image-prompt based generative models

### 2.3. Adapters for multi-view image generation

Methods: - Geometry based (Structure from motion) - Learning based (Diffusion models)  
- fixed camera angles - variable camera angles (SLAM based)

Dzięki technice X [2] uzyskujemy (...). Zgodnie ze standardem bibliografia powinna być uszeregowana alfabetycznie według haseł autorskich, dlatego może się zażyć, że wcześniejsze odwołanie ma wyższą cyfrę.

XYZ używany jest do (...). Dzięki tej technice (...). Powstało wiele rozwinięć tematu takie jak (...) [1, 3, 4, 5, 6], czy (...) [3].

Zauważ, że strona pierwsza rozdziału ma inny styl niż kolejne. Rozdział powinien zawsze rozpoczynać się na nieparzystej stronie. Parzyste strony w nagłówku mają podkreślony wydział: „Faculty of Information and Communication Technology”; nieparzyste (z wyjątkiem pierwszych stron tytułowych): „Wrocław University of Science and Technology”. Jeśli rozdział kończy się na stronie nieparzystej powinna zostać dodana pusta strona bez formatowania, tak aby kolejny rozdział rozpoczął się ponownie od strony nieparzystej. Jakkolwiek staramy się unikać pustych stron.



## 3. Politechnika Wrocławska

Politechnika Wrocławska to państwowa uczelnia techniczna we Wrocławiu. Figure 3.1 przedstawia logo uczelni.



Rysunek 3.1: Logo Politechniki Wrocławskiej

Pamiętaj by nie zostawiać pustych miejsc pod tytułem rozdziału. Należy pamiętać o wprowadzeniu. Dopiero później powinny zostać przedstawione podrozdziały.

### 3.1. Tytuł podrozdziału

Praca może zawierać także tabele, które powinny być czytelne i dobrze opisane. Do wszystkich tabel i rysunków powinny pojawić się odwołania w tekście (oraz komentarze). Podpisy rysunków mają znaleźć się pod rysunkami, a podpisy tabel – nad tabelami. Tabela 3.1 stanowi przykład.

Tabela 3.1: Krótki ale treściwy opis tabeli

| Method | anger        | joy          | optimism     | sadness      |
|--------|--------------|--------------|--------------|--------------|
| model1 | 0.772        | <b>0.751</b> | <b>0.532</b> | 0.673        |
| model2 | 0.727        | 0.661        | 0.307        | 0.629        |
| model3 | 0.761        | 0.739        | 0.498        | <b>0.679</b> |
| model4 | <b>0.782</b> | 0.740        | 0.470        | 0.672        |

Warto także użyć gotowych generatorów do tabel (np. <https://www.tablesgenerator.com/>), niż formatować je samodzielnie. Mistrzem formatowania był Wojtek. Tabela 3.2 o nieco zmienionej treści pochodzi z jego pracy.

Tabela 3.2: Comparison of the compression methods considered for MultiEmo for the sentence level in the mixed domain scenario. The results are averaged on 5 repetitions. The best results are in **bold**.

| Method                 | #Par.      | Mem.<br>[MB]     | Train.<br>[min] | Eval [s]          | Acc.        | F1          | Rec.        | Prec.       |
|------------------------|------------|------------------|-----------------|-------------------|-------------|-------------|-------------|-------------|
| Model <sub>BASE</sub>  | 109M       | 418              | 26.0            | 13.9              | 78.8        | 74.7        | 74.1        | 75.8        |
| Model1                 | 67M        | 255 (1.6x)       | 14.1            | 7.0 (2.0x)        | 77.9        | 73.9        | 73.3        | <b>74.8</b> |
| Model2                 | <b>13M</b> | <b>49 (8.6x)</b> | 6.5             | 2.3 (6.2x)        | 76.7        | 72.4        | 71.7        | 73.9        |
| Model3 <sub>6,TA</sub> | 67M        | 255 (1.6x)       | 14.1            | 7.2 (1.9x)        | 77.7        | 73.8        | 73.5        | 74.3        |
| Model3 <sub>6,TS</sub> | 68M        | 258 (1.6x)       | 62.9            | 7.2 (1.9x)        | <b>78.2</b> | <b>74.4</b> | <b>74.1</b> | 74.7        |
| Model3 <sub>4,TA</sub> | 14M        | 55 (7.6x)        | <b>5.5</b>      | 2.1 (6.6x)        | 76.3        | 72.0        | 71.1        | 73.7        |
| Model3 <sub>4,TS</sub> | 15M        | 56 (7.5x)        | 38.3            | <b>2.0 (6.9x)</b> | 76.3        | 72.2        | 71.5        | 73.4        |
| Model4                 | 67M        | 255 (1.6x)       | 19.1            | 7.8 (1.8x)        | 76.6        | 72.5        | 72.1        | 73.2        |
| Model5                 | 14M        | 55 (7.6x)        | 121.5           | 3.4 (4.1x)        | 77.5        | 72.9        | 72.2        | 74.9        |

### 3.1.1. Tytuł podpodrozdziału

Studia zapewniają podstawową wiedzę z zakresu sztucznej inteligencji i nauki o danych (data science). Rozwijają zarówno umiejętności matematyczne, programistyczne, obliczeniowe, analityczne, jak i umiejętności pracy projektowej w grupie, z nastawieniem na identyfikowanie problemów (naukowych, biznesowych, społecznych) i ich rozwiązywanie z wykorzystaniem technik i metod inteligentnych.

## 3.2. Tytuł podrozdziału 2

Społeczeństwo w swoim rozwoju wchodzi w czwartą rewolucję przemysłową, określaną często terminem „Przemysł 4.0”. Jest on oparty na systemach łączności piątej i wyższych generacji oraz Internecie Rzeczy (IoT), wspieranych sztuczną inteligencją.

Przemysł 4.0 zmienia nie tylko technologie, ale przede wszystkim model biznesu i wymagania stawiane pracownikom. Rewolucja oprze się na danych, które będą zbierane, gromadzone i przetwarzane na każdym etapie prowadzenia biznesu. W firmach wykorzystywane będą nowoczesne technologie, chmury obliczeniowe, wielkie zbiory danych, Internet Rzeczy, rozszerzona rzeczywistość, sztuczna inteligencja czy druk 3D.

Technologia powinna uwzględniać właściwe współdziałanie z człowiekiem. Dlatego należy badać aspekty komunikacji i interakcji człowiek – komputer czy szerzej: człowiek i rozbudowane środowiska robotyczne. Bariery rozwoju Przemysłu 4.0 mają związek przede wszystkim z dostępem do wykształconych kadr. To zrozumiałe w obliczu dużych oczekiwań co do interdyscyplinarnych kompetencji stawianych inżynierom.

Nauka przed wyzwaniem Przyszłość stawia przed europejską nauką trzy kluczowe wyzwania:

- Rozwój innowacji zakłócających – innowacje przełomowe, takie które zmieniają istniejący porządek na rynku, wprowadzając coś zupełnie nowego.
- Biologizacja techniki – zastosowanie praw biologicznych, rządzących procesami zachodzącymi w organizmach żywych, do innych dziedzin.
- Bezpieczeństwo publiczne – nowe rozwiązania jakościowe i innowacje dotyczące: bezpieczeństwa w zakresie IT (m.in. w kontekście ataków cybernetycznych), bezpieczeństwa systemów transportu drogowego, kolejowego i lotniczego, bezpieczeństwa infrastrukturalnego miast i zaopatrzenia w wodę i elektryczność.

Badania i dydaktyka na wysokim poziomie Szczególna rola w przełamywaniu barier rozwoju Przemysłu 4.0 przypadnie uniwersytetom, które w swoich zadaniach statutowych mają: badania naukowe i kształcenie na wysokim poziomie. Zadania te są równie ważne. Nie ma bowiem wysokiej jakości nauczania bez prowadzenia badań naukowych. Nie ma badań naukowych i rozwoju gospodarki bez odpowiednio wykształconej kadry. To uczelnie będą przełamywały główne bariery czwartej rewolucji przemysłowej. Wśród tych uczelni – biorąc pod uwagę kapitał ludzki i społeczny, infrastrukturę badawczą i dydaktyczną – nie może zabraknąć Politechniki Wrocławskiej, aspirującej do bycia uczelnią badawczą. Te istotne zadania w pokonywaniu barier realizują wydziały. Na nie, a przede wszystkim na nowo powstały Wydział Informatyki i Telekomunikacji, spada odpowiedzialność w tej dziedzinie.

Wydział tworzą katedry pochodzące z wydziałów Politechniki Wrocławskiej: Elektroniki, Informatyki i Zarządzania oraz Podstawowych Problemów Techniki. Badania naukowe prowadzone będą w katedrach: Automatyki, Mechatroniki i Systemów Sterowania, Informatyki Technicznej, Systemów i Sieci Komputerowych, Telekomunikacji i Teleinformatyki, Informatyki i Inżynierii Systemów, Informatyki Stosowanej, Inteligencji Obliczeniowej, Podstaw Informatyki.

Oznacza to, że swoim zakresem obejmą większość dziedzin bezpośrednio związanych z Przemysłem 4.0. Realizowane będą zarówno badania podstawowe, jak i wdrożeniowe, skupimy się także na aktywnej współpracy z gospodarką i komercjalizacji wyników. W/w katedry są gwarancją, że badania obejmą większość dziedzin bezpośrednio związanych z Przemysłem 4.0.

Za cel stawiamy sobie przede wszystkim zapewnienie przyjaznych warunków do rozwoju naukowego, zwłaszcza młodym, oraz przejrzystą ścieżkę kariery naukowej. Badania będziemy prowadzić we wszystkich dyscyplinach powiązanych z informatyką.

Kadra ma świadomość, że fundamentem wydziału jest – stojące na wysokim poziomie, nowoczesne, spełniające potrzeby zmieniającego się świata – nauczanie. Odbywać się ono będzie na 12 polsko- i anglojęzycznych kierunkach: Computer Science, Computer Security, Cyberbezpieczeństwo, Informatyczne Systemy Automatyki, Informatyka Algorytmiczna, Informatyka Stosowana, Informatyka Techniczna, Inżynieria Systemów, Sztuczna Inteligencja, Teleinformatyka, Telekomunikacja, Zaufane Systemy Sztucznej Inteligencji. Lista kierunków będzie sukcesywnie modyfikowana w odpowiedzi na wyzwania stawiane przez uwarunkowania społeczno-gospodarcze.



## 4. Conclusions and Future

Zakończenie, podsumowuje najważniejsze wnioski, podaje możliwości dalszego rozwinięcia wykonanych prac i wskazuje obszar potencjalnego zastosowania pracy. Rezultaty pracy mają charakter poznawczy, mogą mieć charakter użytkowy. Należy dokonać analizy uzyskanych wyników. Rezultaty powinny charakteryzować się oryginalnością, a nawet w pewnym stopniu nowatorstwem. Praca zawiera (...). Zostało pokazane (...). Eksperymenty wykazały (...). Tu piszemy wnioski i obserwacje.

Widzimy, że (...). Z tego powodu przyszła praca powinna obejmować (...).





# Bibliografia

- [1] R. Gao, A. Holynski, P. Henzler, A. Brussee, R. Martin-Brualla, P. Srinivasan, J. T. Barron, and B. Poole. Cat3d: Create anything in 3d with multi-view diffusion models. *arXiv preprint arXiv:2405.10314*, 2024.
- [2] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2 edition, 2003.
- [3] Z. Huang, Y. Guo, H. Wang, R. Yi, L. Ma, Y.-P. Cao, and L. Sheng. Mv-adapter: Multi-view consistent image generation made easy. *arXiv preprint arXiv:2412.03632*, 2024.
- [4] R. Liu, R. Wu, B. V. Hoorick, P. Tokmakov, S. Zakharov, and C. Vondrick. Zero-1-to-3: Zero-shot one image to 3d object, 2023.
- [5] D. Watson, W. Chan, R. Martin-Brualla, J. Ho, A. Tagliasacchi, and M. Norouzi. Novel view synthesis with diffusion models, 2022.
- [6] W. Yu, J. Xing, L. Yuan, W. Hu, X. Li, Z. Huang, X. Gao, T.-T. Wong, Y. Shan, and Y. Tian. Viewcrafter: Taming video diffusion models for high-fidelity novel view synthesis, 2024.

# Spis rysunków

|  |   |
|--|---|
| 3.1 Logo Politechniki Wrocławskiej . . . . . | 5 |
|--|---|

# Spis tabel

|     |   |   |
|-----|---|---|
| 3.1 | Krótki ale treściwy opis tabeli . . . . .   | 5 |
| 3.2 | Comparison of the compression methods considered for MultiEmo for the sentence level in the mixed domain scenario. The results are averaged on 5 repetitions. The best results are in <b>bold</b> . . . . . | 6 |