Logo

Description automatically generated

**Damian Andrysiak**

Kierunek: informatyka

Specjalność: informatyka stosowana

Numer albumu: 374341

**Badanie efektywności zwiększania rozdzielczości obrazów za pomocą sieci neuronowych**

Praca magisterska

wykonana pod kierunkiem

dr. hab. prof. nadzw. UŁ.

Pawła Kowalczyka

w Katedrze Fizyki Ciała Stałego

WFiIS UŁ

**Łódź 2023**

# Spis treści

[Spis treści 2](#_Toc131968162)

[1. Wstęp 3](#_Toc131968163)

[2. Zbiór danych 5](#_Toc131968164)

[2.1. Proces generowania danych 5](#_Toc131968165)

[2.2. Przygotowanie silnika do generowania danych 9](#_Toc131968166)

[2.3. Przygotowanie projektu do generowania danych 9](#_Toc131968167)

[2.4. Normalizacja danych 9](#_Toc131968168)

[3. Podsumowanie i bibliografia 9](#_Toc131968169)

[3.1. Podsumowanie 9](#_Toc131968170)

[3.2. Bibliografia 9](#_Toc131968171)

# Wstęp

Podstawowym celem pracy pt. „Badanie efektywności zwiększania rozdzielczości obrazów z wykorzystaniem sieci neuronowych” będzie zbadanie możliwości powiększania obrazów HDR z rozdzielczości 1080p do rozdzielczości 4k (tj. 3840x2160 pikseli) z wykorzystaniem sieci neuronowych.

Na samym wstępie, warto wspomnieć o technice, dzięki której można uzyskać zwiększenie rozdzielczości zdjęć, jest nią SISR (eng. Single Image Super Resolution - Uzyskanie Super Rozdzielczości z Pojedynczego Obrazu). Metoda ta wykorzystuje głębokie sieci konwolucyjne do uzyskania lepszej jakości oraz rozdzielczości z obrazu o niższej rozdzielczości w odróżnieniu od metod czysto analitycznych takich jak interpolacja najbliższych sąsiadów (eng. Nearest-neighbors), bilionowa (eng. Bilinear) lub bikubiczna (eng. Bicubic). Sieci neuronowe są w stanie nauczyć się bardziej złożonych zależności miedzy pikselami przez co generowane obrazy są wyższej jakości i zawierają więcej szczegółów. Ich popularność rośnie z roku na rok miedzy innymi dzięki ich możliwościom adaptacji do wielu problemów, dostępności dużej ilości wytrenowanych ówcześnie modeli, które są gotowe do użycia oraz coraz większej ilości API (eng. Application Programming Interface), które umożliwiają tworzenie własnych modeli w oparciu o gotowe dokumentacje, przystępne nawet dla osób nieskorelowanych ze sztuczna inteligencja.

W opisywanej pracy magisterskiej została użyta zmodyfikowana architektura głębokiej sieci neuronowej typu UNET[1]. Składa się z ona trzech głównych bloków, jakimi są kolejno blok zmniejszający rozdzielczość (eng. Downsample Block/Encoder), który jest odpowiedzialny za ekstrakcję cech oraz zmniejszanie rozdzielczości, tudzież kompresję obrazu, pozwalająca na wydobywanie szczegółów, następnie występuje wąskie gardło (eng. Bottleneck), który odpowiada za ostateczną ekstrakcję cech, a ostatnim blokiem jest blok zwiększający rozdzielczość (eng. Upsample Block/Decoder), przynajmniej do rozdzielczości wejściowej obrazu. Cześć pracy, związana bezpośrednio z siecią neuronowa została napisana z użyciem biblioteki PyTorch.

Kolejnym ważnym aspektem w pracy jest proces generowania danych. Do tego celu został wykorzystany silnik graficzny. Silnikiem graficznym użytym w pracy jest Unreal Engine w wersji 4.26.2, który został poddany modyfikacjom umożliwiającym zapisywanie na dysk pojedynczych klatek generowanych przez silnik podczas używania sekwencji w danym projekcie. Unreal Engine jest to jeden z najbardziej, jak nie najbardziej popularny silnik graficzny służący do renderowania wysokiej jakości danych oraz tworzenia gier komputerowych.

Opisywana praca magisterska została podzielona na dwa podstawowe pod problemy: Silnik graficzny umożliwiający generowanie zbioru danych oraz sieć neuronowa wraz z całokształtem procesu uczenia oraz walidacji wyników. Oba te procesy zostały opisane ze szczegółami w rozdziale nr 2.

# Zbiór danych

## Proces generowania danych

Proces generowania danych jest ścisłe związany z wspomnianym w rozdziale 1 silnikiem o nazwie Unreal Engine w wersji 4.26.2. Silnik ten posiada bardzo rozbudowany proces generowania obrazów, które z wersji na wersje zaczynają być nierozróżnialne, kiedy to porównamy obrazy wygenerowane przez silnik a rzeczywiste zdjęcia. Na przykładzie [Rysunek 2.1] możemy wyróżnić sześć podstawowych potoków graficznych (eng. Graphic pipelines), które maja miejsce podczas generowania pojedynczej klatki (wykluczając dwa ostatnie potoki tj. eng. User Interface oraz eng. Backbuffer Resolution):

1. **Geometria widoku (eng. View Geometry)** jest to proces w potoku graficznych, który służy do przygotowania macierzy projekcji, translacji itp. Dla aktualnej ramki/klatki oraz:
   1. **Tworzenia geometrii/bufora głębokości (eng. Depth Buffer)** odpowiedzialnego za informacje o odległości położenia obiektu od kamery w wartościach składowej Z sceny trójwymiarowej.
   2. **Tworzenia geometrii/bufora prędkości/ruchu (eng. Velocity Buffer)** odpowiedzialnego za informacje o poruszaniu się obiektu w wartościach X oraz Y sceny trójwymiarowej.
2. **Głębokość pola widoku (eng. Depth of Field)** jest odpowiedzialny za efekt rozmycia obiektów na scenie bazując na wartościach z bufora głębokości. Obiekty, na których obecnie gracz nie skupia swojej uwagi, staja się rozmyte, jak pokazano na [Rysunek 2.2].
3. **Metoda anty-aliasingowa (eng. Anti-Aliasing Method)** jest odpowiedzialna za wygładzanie krawędzi, jak pokazano na [Rysunek 2.3] przez wygładzeniem krawędzi oraz po ich wygładzeniu [Rysunek 2.4]. Najbardziej popularne metody Anty-Aliasingowe to FXAA (eng. Fast Aproximate Anti-Aliasing), MSAA (eng. Multi-Sampling Anti-Aliasing), TAA[2] (eng. Temporal Anti-Aliasing) oraz SSAA (Super-Sampling Anti-Aliasing).
4. **Metoda rozmywania ruchu kamery (eng. Motion Blur)** polega na dodaniu do sceny efektu rozmycia podczas ruchu kamery [Rysunek 2.5].
5. **Efekt rozmycia światła (eng. Bloom)** jest stosowany do uzyskania efektu rozmycia/poświaty ze źródła światła [Rysunek 2.6].
6. **Mapowanie tonów (eng. Tonemapping/Tonemapper)** służy do mapowania wysokiego zakresu kolorow (eng. HDR – High Dynamic Range) do niskiego zakresu kolorów (eng. LDR – Low Dynamic Range).

Table, timeline

Description automatically generatedRysunek . Przykładowy wygląd procesów graficznych potrzebnych do wygenerowania pojedynczej klatki w silniku Unreal Engine.

A picture containing indoor, engine

Description automatically generatedRysunek . Przykładowy wygląd sceny po aplikacji procesu głębokości pola widoku/sceny (eng. Depth of Field), aktualny obiekt (A), na którym skupia się gracz pozostaje w wysokiej ostrości, reszta obiektów (np. B), staje się rozmyta.Chart, line chart

Description automatically generated

Rysunek . Wygląd krawędzi przed ich wygładzeniem metoda Anty-Aliasingowa.

Chart, line chart

Description automatically generated

Rysunek . Wygląd krawędzi po ich wygładzeniu metodą Anty-Aliasingowa.

A picture containing indoor, music

Description automatically generated

Rysunek . Ramka po aplikacji efektu rozmycia zależnego od ruchu kamery.





Rysunek . Obrazek A przedstawia scenę przed efektem rozmycia światła, a obrazek B tuz po jego zastosowaniu.

## Przygotowanie silnika do generowania danych

ADASDDDDADASD

## Przygotowanie projektu do generowania danych

DADASDASDADADAD

## Normalizacja danych

ASDADASDADASD

# Podsumowanie i bibliografia

## Podsumowanie

ADDDASD

## Bibliografia

[1] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2015. doi: 10.1007/978-3-319-24574-4\_28.

[2] L. Yang, S. Liu, and M. Salvi, “Number 2 STAR-State of The Art Report,” 2020. [Online]. Available: http://onlinelibrary.wiley.com.Thisarticlemaybeusedfornon-commercialpurposesinaccordancewiththeWileySelf-ArchivingPolicy[http://olabout.wiley.com/WileyCDA/Section/id-820227.html].