**Politechnika Świętokrzyska**

w Kielcach

Wydział Elektrotechniki, Automatyki i Informatyki

Mateusz Gajos

Numer albumu: 091209,b

Aplikacja webowa wspomagająca pracę fotografa z zastosowaniem GAN

Praca dyplomowa inżynierska

Na kierunku Teleinformatyka

Opiekun pracy dyplomowej:

Dr inż. Katarzyna Rutczyńska-Wdowiak

Kielce, 2024

Składam serdeczne podziękowania

XXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXX

Tu miejsce na Zadanie z dziekanatu

Tu miejsce na oświadczenie

**Aplikacja webowa wspomagająca pracę fotografa z zastosowaniem GAN**

**Streszczenie**

Zasadniczym założeniem w niniejszej pracy jest opracowanie aplikacji webowej, której celem jest usprawnienie procesu pracy fotografów poprzez wykorzystanie technologii GAN (Generatywnych Sieci Współzawodniczących, eng: Generative Adversarial Network). Głównym celem jest stworzenie platformy która pozwala na edytowanie oraz generowanie nowych obrazów na podstawie istniejących fotografii, co może znacznie ułatwić bądź zautomatyzować proces obróbki zdjęć.

W ramach pracy przeprowadzono analizę problemu, identyfikujący wyzwania i potrzeby współczesnych fotografów w zakresie postprodukcji zdjęć. Następnie, oparto się na teorii GAN, aby zrozumieć i zaimplementować algorytmy generowania obrazów. Kluczowym elementem aplikacji jest umożliwienie użytkownikowi interakcyjnego dostępu do narzędzi opartych na sztucznej inteligencji, umożliwiających szybką i efektywną edycję zdjęć.

Dodatkowo, w pracy skoncentrowano się na aspektach praktycznych, takich jak projektowanie intuicyjnego interfejsu użytkownika, optymalizacja wydajności aplikacji webowej oraz zastosowanie odpowiednich technologii do implementacji funkcji opartych na GAN. W trakcie implementacji przeprowadzono testy aplikacji, oceniając jej skuteczność i użyteczność.

**Słowa kluczowe**

Fotografia

Postprodukcja zdjęć

Generatywne sieci współzawodniczące (GAN)

Aplikacja webowa

Django

**Spis treści**

1 Wprowadzenie

2 Omówienie działania oraz problemów implementacji generatywnych sieci współzawodniczących

3 Przegląd rynku z zakresu istniejących rozwiązań alternatywnych problemów

4 Założenia i struktura programu komputerowego

5 Charakterystyka wykorzystanych narzędzi programistycznych

6 Omówienie funkcjonalności programu

7 Testy programu oraz analiza uzyskanych wyników

8 Podsumowanie

9 Literatura

10 Załączniki

**Wprowadzenie**

Niniejsza praca inżynierska podejmuje problematykę przedstawienia, zaprojektowania oraz stworzenia innowacyjnej aplikacji webowej, której celem jest wspomaganie pracy fotografa. Jednak, czynnikiem wyróżniającym tą aplikację poza standardowe narzędzia jest zastosowanie Generatywnych Sieci Współzawodniczących (GAN). GAN to rodzaj sztucznej inteligencji, która jest zdolna do generowania nowych treści, w tym obrazów oraz filmów, poprzez rywalizację między dwoma sieciami neuronowymi. W kontekście naszej aplikacji, GANy zostaną wykorzystane do stworzenia nowatorskich narzędzi wspomagających proces tworzenia i edycji fotografii.

W dzisiejszym dynamicznym środowisku technologicznym, rozwój technologii stanowi kluczowy element poprawy efektywności i wydajności w wielu dziedzinach życia zawodowego. Jednym z obszarów, który znacząco zyskał na znaczeniu w erze cyfrowej, jest fotografia. Fotografowie, zarówno profesjonaliści, jak i amatorzy, korzystają z zaawansowanych narzędzi do tworzenia, edytowania i udostępniania swoich dzieł. W tym znaczeniu, aplikacje webowe stanowią kluczowy element ułatwiający pracę fotografa, umożliwiający im efektywne zarządzanie projektem, archiwizację zdjęć oraz współpracę z klientem.

Aplikacja w pełni bazuje na dwóch głównych technologiach: Django oraz Python. Django, popularny framework do tworzenia aplikacji webowych, umożliwia szybkie i efektywne budowanie aplikacji o zaawansowanych funkcjonalnościach. Python, jako język programowania, dostarcza potężnych narzędzi do implementacji algorytmów sztucznej inteligencji, co czyni go idealnym wyborem dla projektu, który łączy te dwie dziedziny.

Celem tej pracy jest nie tylko stworzenie aplikacji webowej, ale także eksploracja potencjału, jaki tkwi w zastosowaniach GAN’ów w dziedzinie generowania obrazów. W dalszej części tej pracy zostaną szczegółowo omówione etapy projektowania, implementacji oraz testowania aplikacji, a także analiza wyników uzyskanych dzięki wykorzystaniu GAN. Ostateczny rezultat ma przyczynić się do rozwoju narzędzi wspomagających pracę fotografa, jednocześnie poszerzając horyzont w zakresie możliwości, jakie oferuje sztuczna inteligencja

**Omówienie działania oraz problemów implementacji generatywnych sieci współzawodniczących**

2.1 Wyjaśnienie terminologii

Generatywne Sieci Współzawodniczące to rodzaj sztucznych sieci neuronowych składających się z dwóch głównych komponentów: generatora i dyskryminatora. GAN zostały wprowadzone przez Iana Goodfellowa i jego współpracowników w 2014 roku i od tego czasu zyskały ogromną popularność w dziedzinie uczenia maszynowego, głównie w dziedzinie generacji obrazów, ale także w innych dziedzinach, takich jak przetwarzanie języka naturalnego.

2.2 Omównienie działania GAN

A) Generatory, Zadaniem generatora jest tworzenie nowych przykładów danych, które są trudne do odróżnienia od prawdziwych danych. Na wejściu otrzymuje losowy szum (wektor zmiennych losowych) i generuje przykłady danych.

B) Dyskryminator. Zadaniem dyskryminatora jest ocenianie, czy dostarczony mu przykład danych jest prawdziwy (pochodzi z rzeczywistego zbioru danych) czy sztuczny (wygenerowany przez generator). Na wejściu otrzymuje zarówno prawdziwe, jaki i wygenerowane dane i stara się je rozróżnić.

C) Trening: Generator i dyskryminator są trenowane wspólnie w procesie adversarialnym. Podczas treningu generator stara się oszukać dyskryminator, a dyskryminator stara się nauczyć odróżniać prawdziwe dane od wygenerowanych. Proces ten prowadzi do doskonalenia zarówno generatora, jak i dyskryminatora.

D) Equilibrium: Optymalne działanie GAN osiągane jest, gdy generator tworzy dane trudne do odróżnienia od prawdziwych, a dyskryminator jest w stanie je rozróżnić.

2.3 Problemy implementacyjne GAN:

A) Mode Collapse: Generator może nauczyć się produkować tylko jedną, powtarzającą się kategorię danych, ignorując inne.

B) Niestabilność treningu: Proces treningu GAN jest zazwyczaj niestabilny. Mogą wystąpić oscylacje w jakości generowanych danych i wyniki mogą być trudne do kontrolowania.

C) Dysproporcja w mocy dyskryminacyjnej: Jeśli dyskryminator jest zbyt silny, może dominować nad generatorem, uniemożliwiając mu generowanie realistycznych danych.

D) Dobieranie hiperparametrów: Wybór odpowiednich hiperparametrów dla GAN może być trudnym i czasochłonnym procesem.

E) Overfitting: GAN podatny jest na overfitting, zwłaszcza jeżeli zestaw danych treningowych jest mały.

F) Trudności z oceną jakości: Brak jednoznacznej metryki oceny jakości generowanych danych sprawia, że trudno jest ocenić, kiedy model jest wystarczająco dobrze wytrenowany.

W związku z tym, implementacja GAN wymaga eksperymentacji, doboru hiperparametrów i podejścia do radzenia sobie z problemami specyficznymi dla danego zadania. Wiele nowych technik i architektury zostało zaproponowanych w celu poprawy stabilności treningu i jakości generowanych danych.

**Przegląd rynku z zakresu istniejących rozwiązań alternatywnych problemów**

Przedmiotem niniejszego podrozdziału jest dokładny przegląd rynku istniejących rozwiązań alternatywnych, które konkurują lub współpracują z propozycją naszej aplikacji webowej wspomagającej pracę fotografa z zastosowaniem GAN. Wnikliwe zrozumienie obecnych rozwiązań na rynku jest kluczowe dla skutecznego dostosowania naszej aplikacji do potrzeb użytkowników oraz identyfikacji unikalnych cech, które wyróżnią naszą propozycję.

Analiza konkurencji obejmuje zarówno szeroko dostępne narzędzia jak i specjalistyczne rozwiązania, które mogą obejmować różne aspekty procesu fotograficznego, takie jak retusz, generowanie efektów wizualnych, czy automatyzacja prac związanych z edycją zdjęć. Wszystko to z myślą o dostarczeniu kompleksowego spojrzenia na obecny stan technologii związanych z fotografią, co pozwoli na trafniejsze zdefiniowanie naszych celów projektowych.

Ponadto, analiza konkurencji pozwoli zidentyfikować luki na rynku oraz potencjalne obszary, w których nasza aplikacja może się wyróżniać. Skupimy się na różnych aspektach, takich jak użyteczność, dostępność, funkcjonalność, elastyczność i jakość efektów generowanych przez rozważane rozwiązania.

W dalszej części tego podrozdziału przedstawione zostaną kluczowe konkurencyjne rozwiązania, z ich charakterystyką, mocnymi i słabymi stronami, a także analizą, jakie wyzwania stawiają przed użytkownikami. Ostatecznym celem jest stworzenie solidnej podstawy dla naszej aplikacji, która nie tylko skutecznie konkurować będzie z obecnymi rozwiązaniami, ale także przekroczy oczekiwania użytkowników poprzez nowoczesność, innowacyjność i zaawansowane wykorzystanie sztucznej inteligencji.

a. Adobe Lightroom i Adobe Photoshop:

Opis: Adobe Lightroom i Photoshop są jednymi z najpopularniejszych narzędzi wśród profesjonalistów ds. obróbki zdjęć. Oferują szeroką gamę funkcji, umożliwiając retusz, korektę kolorów, edycję RAW, oraz inne zaawansowane techniki. Lightroom jest skoncentrowany na organizacji i edycji zdjęć, podczas gdy Photoshop oferuje bardziej zaawansowane funkcje retuszu.

b. Capture One:

Opis: Capture One to kolejne profesjonalne oprogramowanie do obróbki zdjęć, które jest szczególnie popularne wśród fotografów zajmujących się fotografią studyjną. Zapewnia zaawansowane narzędzia do edycji, obsługę plików RAW, oraz intuicyjny interfejs.

c. Darktable:

Opis: Darktable to bezpłatne i otwarte oprogramowanie do obróbki zdjęć, dedykowane głównie systemom operacyjnym Linux. Choć nie jest tak wszechstronne jak niektóre płatne narzędzia, to stanowi dobrą alternatywę dla budżetowych fotografów.

Rynek aplikacji wspomagających pracę fotografa z zastosowaniem GAN:

a. DeepArt:

Opis: DeepArt wykorzystuje sztuczną inteligencję, w tym GAN, do transformacji zdjęć w dzieła sztuki, pozwalając fotografom na eksperymentowanie z różnymi stylami artystycznymi.

Zalety: Unikalne efekty sztuki, łatwość użycia.

Wady: Ograniczone do transformacji stylów, mniej narzędzi edycyjnych.

b. DALL-E by OpenAI:

Opis: DALL-E to model GAN stworzony przez OpenAI, zdolny do generowania kreatywnych obrazów na podstawie podanego opisu tekstowego.

Zalety: Kreacja obrazów na podstawie opisu, potencjał tworzenia unikalnych zdjęć.

Wady: Wczesne stadium rozwoju, bardziej skomplikowane w użyciu.

c. Let's Enhance:

Opis: Platforma używająca GAN do poprawy jakości i rozdzielczości zdjęć, co może być przydatne dla fotografów chcących poprawić starą fotografię.

Zalety: Poprawa jakości zdjęć, łatwe w użyciu.

Wady: Ograniczone do poprawy jakości, mniej narzędzi edycyjnych.

Przy analizie tych platform, ważne jest zwrócenie uwagi na kombinację narzędzi edycyjnych, łatwość użycia, dostępność funkcji GAN oraz reakcje społeczności użytkowników. Ostateczny wybór zależy od indywidualnych preferencji i potrzeb fotografa.b. Artbreeder:

Opis: Artbreeder to platforma, która wykorzystuje technologię GAN do tworzenia sztuki poprzez mieszanie różnych obrazów. Fotografowie mogą korzystać z niej do eksperymentowania z różnymi stylami i efektami w celu stworzenia unikalnych kompozycji.

c. GANPaint Studio:

Opis: GANPaint Studio to narzędzie oparte na technologii GAN, które umożliwia manipulację elementami obrazu, takimi jak dodawanie, usuwanie lub modyfikacja obiektów na zdjęciu. To może być przydatne dla fotografów, którzy chcą dostosować sceny na zdjęciach.

**Założenia i struktura programu komputerowego**

**Charakterystyka wykorzystanych narzędzi programistycznych**

Python:

Zastosowanie: Python to interpretowany, wysokopoziomowy język programowania, wykorzystywany w szerokim zakresie dziedzin, od analizy danych po sztuczną inteligencję. Jego składnia jest czytelna i przejrzysta, co sprzyja szybkiemu prototypowaniu i skracaniu czasu tworzenia rozwiązań.

Django:

Zastosowanie: Django jest pełnoprawnym frameworkiem do tworzenia aplikacji webowych, opartym na wzorcu architektonicznym Model-Widok-Kontroler (MVC). Posiada wbudowane narzędzia do obsługi baz danych, zarządzania sesją, a także system autentykacji, co przyspiesza rozwój nowoczesnych aplikacji internetowych.

Numpy:

Zastosowanie: Numpy jest potężną biblioteką do obliczeń numerycznych, szczególnie przydatną w kontekście przetwarzania danych numerycznych. Udostępnia efektywne struktury danych, takie jak tablice wielowymiarowe, oraz funkcje matematyczne do operacji na danych.

Torch:

Zastosowanie: Torch to framework do uczenia maszynowego i głębokiego uczenia, który znalazł szerokie zastosowanie w badaniach nad sztuczną inteligencją. Dzięki elastycznej architekturze, jest używany do implementacji i szkolenia skomplikowanych modeli.

Bootstrap:

Zastosowanie: Bootstrap to responsywny framework front-endowy, oparty na HTML, CSS i JavaScript. Zapewnia gotowe komponenty interfejsu użytkownika, co pozwala na szybkie projektowanie estetycznych i funkcjonalnych stron internetowych.

Visual Studio Code:

Zastosowanie: Visual Studio Code to zaawansowane i lekkie środowisko programistyczne, które wspiera wiele języków programowania, w tym Pythona. Oferuje funkcje takie jak inteligentne podpowiedzi, debugger, a także integrację z systemem kontroli wersji.

Pillow:

Zastosowanie: Pillow to rozwinięcie oryginalnej biblioteki Python Imaging Library (PIL) i stanowi potężne narzędzie do przetwarzania obrazów. Obsługuje różne formaty plików obrazów i zapewnia funkcje takie jak przycinanie, skalowanie i konwersje kolorów.

CV2 (OpenCV):

Zastosowanie: OpenCV to biblioteka do komputerowego przetwarzania obrazów i analizy wideo. Jest powszechnie stosowana w projektach związanych z rozpoznawaniem obrazów, detekcją obiektów i przetwarzaniem wizualnym.

Git:

Zastosowanie: Git to system kontroli wersji, który ułatwia śledzenie i zarządzanie zmianami w kodzie źródłowym. Dzięki niemu można skutecznie pracować w zespole, śledzić historię projektu i łatwo przywracać wcześniejsze wersje.

GitHub:

Zastosowanie: GitHub to platforma hostingowa dla projektów opartych na Git, która umożliwia łatwe udostępnianie kodu, zarządzanie zadaniem i współpracę z innymi programistami. Jest niezastąpionym narzędziem dla pracy zespołowej.

Hugging Face:

Zastosowanie: Hugging Face to społeczność i platforma, która skupia się na udostępnianiu modeli i zestawów danych do uczenia maszynowego, szczególnie w dziedzinie przetwarzania języka naturalnego. Jest cennym źródłem pre-trenowanych modeli i narzędzi NLP.

HTML (HyperText Markup Language):

HTML to język znaczników używany do tworzenia struktur i hierarchii treści na stronach internetowych. Jest to podstawowy składnik każdej witryny internetowej i stanowi szkielet dokumentu, definiując strukturę informacji. Oto kluczowe elementy HTML:

Tagi:

HTML składa się z zestawu tagów, które otaczają różne elementy treści. Na przykład, <head> zawiera metadane, a <body> zawiera główną treść strony.

Atrybuty:

Tagi mogą zawierać atrybuty, które dostarczają dodatkowych informacji o elemencie. Na przykład, <img> może mieć atrybuty takie jak "src" (źródło obrazu) i "alt" (alternatywny tekst).

Semantyka:

HTML wprowadza semantykę, co oznacza, że nie tylko definiuje strukturę, ale także znaczenie poszczególnych elementów. Na przykład, <header>, <nav>, <article> i <footer> określają znaczenie obszarów na stronie.

Formularze:

HTML zawiera elementy do tworzenia interaktywnych formularzy, takich jak <form>, <input>, <select> i <button>. Pozwalają one na interakcję użytkownika, zbieranie danych i przesyłanie ich na serwer.

Osadzanie multimediów:

Elementy takie jak <img>, <audio> i <video> pozwalają na osadzanie multimediów w stronach internetowych, co jest kluczowe dla bogatych treści.

HTML jest fundamentem struktury witryny, a jego poprawne użycie ma kluczowe znaczenie dla dostępności, SEO (optymalizacji pod kątem wyszukiwarek internetowych) oraz ogólnej użyteczności strony.

CSS (Cascading Style Sheets):

CSS to język arkuszy stylów używany do stylizacji elementów HTML, nadawania im wyglądu i układu. Oto główne aspekty CSS:

Selektory:

CSS korzysta z selektorów, aby określić, które elementy HTML mają być stylizowane. Na przykład, selektor p odnosi się do wszystkich paragrafów, a #header odnosi się do elementu o identyfikatorze "header".

Właściwości i Wartości:

CSS definiuje różne właściwości (np. color, font-size, margin) i wartości (np. red, 12px, 20px) dla stylizowanych elementów. Właściwości te kontrolują wygląd i układ elementów.

Kaskadowość:

Kaskadowość w CSS oznacza, że stylizacja może być dziedziczona, a styl jednego elementu może być zastępowany przez stylizację elementu nadrzędnego. To zapewnia spójność i składność wizualną witryny.

Pseudoklasy i Pseudoelementy:

CSS umożliwia stosowanie stylów do elementów w określonych sytuacjach za pomocą pseudoklas i pseudoelementów. Na przykład, :hover aktywuje styl po najechaniu myszką na element.

Responsywność:

Media queries pozwalają na dostosowywanie stylów do różnych urządzeń, co umożliwia projektowanie responsywnych stron internetowych, dostosowujących się do różnych rozmiarów ekranów.

CSS jest kluczowy do tworzenia estetycznych i funkcjonalnych projektów, a także zapewnia spójność wizualną na całej stronie.

JavaScript (JS):

JavaScript to skryptowy język programowania, który umożliwia dynamiczną interakcję użytkownika na stronie internetowej. Poniżej znajdziesz kluczowe aspekty JavaScript:

Zdarzenia:

JavaScript reaguje na zdarzenia takie jak kliknięcia, najechania myszką, wciskanie klawiszy itp. Zdarzenia te umożliwiają interakcję użytkownika z witryną.

Manipulacja DOM:

JavaScript pozwala na dynamiczną manipulację Drzewem Obiektów Dokumentu (DOM), czyli reprezentacją struktury HTML. To umożliwia dodawanie, usuwanie i modyfikowanie elementów strony.

Funkcje i Zmienne:

Skrypty JavaScript zawierają funkcje i zmienne, które umożliwiają strukturyzowanie kodu, przechowywanie danych i definiowanie operacji.

AJAX (Asynchronous JavaScript and XML):

AJAX umożliwia asynchroniczne ładowanie danych z serwera, co pozwala na aktualizację zawartości strony bez konieczności jej przeładowywania. To kluczowe dla responsywności aplikacji.

Obiekty i Klasy:

JavaScript obsługuje obiekty i klasy, co pozwala na programowanie obiektowe. Działa to zgodnie z paradygmatem programowania, w którym programy są modelowane jako zbiory obiektów, z którymi można manipulować.

JavaScript jest niezbędny dla tworzenia interaktywnych i dynamicznych stron internetowych, co zwiększa zaangażowanie użytkownika i poprawia doświadczenie przeglądania.

AJAX (Asynchronous JavaScript and XML):

AJAX to technika programowania, która umożliwia asynchroniczne wysyłanie i odbieranie danych z serwera, bez konieczności odświeżania całej strony. Kluczowe cechy AJAX:

Asynchroniczność:

AJAX pozwala na wysyłanie żądań do serwera i odbieranie odpowiedzi bez zatrzymywania wykonywania innych operacji na stronie. To zapewnia płynne i responsywne interakcje.

XMLHttpRequest:

Tradycyjnie, AJAX używa obiektu XMLHttpRequest do komunikacji z serwerem. Jednak współcześnie stosuje się również nowsze API takie jak fetch.

Dynamika Strony:

AJAX pozwala na dynamiczną aktualizację zawartości strony bez konieczności przeładowywania. Można dzięki temu ładować jedynie niezbędne dane, co przyspiesza działanie witryny.

Formaty Danych:

Choć początkowo AJAX wykorzystywał XML, obecnie najczęściej stosuje się formaty danych takie jak JSON. Jest bardziej czytelny i łatwy do przetwarzania w języku JavaScript.

Frameworki i Biblioteki:

Współcześnie wiele frameworków i bibliotek (takich jak jQuery, axios) ułatwia implementację AJAX, eliminując część złożoności i zapewniając bardziej czytelny kod.

AJAX jest nieodłączną częścią interaktywnych aplikacji internetowych, pozwalając na szybką wymianę danych z serwerem i tworząc bardziej dynamiczne i atrakcyjne wrażenia dla użytkowników.

**Omówienie funkcjonalności programu**

Cała aplikacja została podzielona na 4 główne segmenty:

System Bazodanowy ( zapis danych do bazy danych)

System Sesji ( Logowanie, rejestracja, autoryzacja)

Aplikacja do zarządzania Zdjęciami, Tworzenie sesji użytkownika, edycja zdjęć, dodawania zdjęć, komunikacja klienta z fotografem

Aplikacja do zarządzania GAN, Upscaling, Generowanie obrazów z tekstu, Generowanie obrazów ze zdjęcia, Generowanie obrazów z tekstu oraz zdjęcia, Edycja obrazów Inpaint za pośrednictwem maski