

NOTATKA 1:

A review on Generative Adversarial Networks for image generation

Notka bibliograficzna:

V. L. T. de Souza, B. A. D. Marques, H. C. Batagelo, J. P. Gois, "A review on Generative Adversarial Networks for image generation", Computers & Graphics, Elsevier Ltd., 2023, DOI:10.1016/j.cag.2023.05.010

Omówienie artykułu:

Poruszony problem inżynierski, cel badawczy:

Celem badawczym jest przedstawienie przeglądu podstawowych zasad różnorodnych architektur Generative Adversarial Networks (GANs), koncentrując się na ich zastosowaniach w generowaniu obrazów.

Formalizacja rozwiązania, opis modelu lub adaptacji algorytmu:

GANy są architekturą deep learningu wykorzystującą dwie sieci – generator (G) i dyskryminator (D) – które konkurują ze sobą, aby stworzyć realistyczne, ale wcześniej niewidziane próbki. Dyskryminator jest szkolony do odróżniania instancji syntetycznych stworzonych przez generator od danych rzeczywistych, podczas gdy generator dąży do produkcji bardzo realistycznych próbek. W artykule omówiono różne modyfikacje architektury, takie jak Deep Convolutional GAN (DCGAN), Conditional GAN (CGAN), StyleGAN, CycleGAN oraz Wasserstein GAN (WGAN).

Dane wejściowe (forma, struktura):

W podstawowym GAN wejściem jest wektor szumu z próbkowany losowo z rozkładu zwanego przestrzenią utajoną Z . W przypadku CGAN do wejść generatora i dyskryminatora dołączany jest warunek y , np. klasa danych, wektor instrukcji, inne obrazy lub osadzone zdanie tekstowe.

Przebieg algorytmu:

Kluczową zasadą jest trening adwersarialny. Generator G tworzy syntetyczny wynik $G(z)$, który jest prezentowany dyskryminatorowi D . D klasyfikuje próbki rzeczywiste i fałszywe $G(z)$. Dyskryminator jest trenowany, aby maksymalizować funkcję celu $V(D, G)$ (gdzie $D(x)=1$ i $D(G(z))=0$), podczas gdy generator jest trenowany, aby minimalizować tę funkcję, dążąc do wyprodukowania próbek, które oszukają dyskryminatora. W praktyce funkcja celu jest często rozdzielana, aby zapobiec nasyceniu gradientu generatora.

Rezultat (Zastosowanie Przemysłowe/Biznesowe):

W obszarze przetwarzania obrazu kluczowe aplikacje to: synteza obrazu, manipulacja obrazem z przestrzeni utajonej, tłumaczenie obraz-obraz (image-to-image translation), transfer stylu, super-rozdzielczość i naprawa obrazu.

GANy mają też znaczące zastosowanie przemysłowe i badawcze w przetwarzaniu obrazu wideo, medycynie, odkrywaniu leków i chemii, meteorologii, rynku akcji, oraz geologii.

NOTATKA 2:

A Review on the Human Face Generation Approaches

Notka bibliograficzna: R. Ravinder Reddy, D. Raman, Saphalya Peta, Vineela Bellamkonda, "A Review on the Human Face Generation Approaches", International Conference on Inventive Computation Technologies (ICICT), 2024, DOI:10.1109/ICICT60155.2024.10544743

Omówienie artykułu:

Poruszony problem inżynierski, cel badawczy:

Badanie ma na celu wykorzystanie Conditional Generative Adversarial Networks (CGANs) do syntezy ludzkich twarzy, skupiając się na zachowaniu tożsamości (identity preservation) oraz kontrolowanej generacji w oparciu o atrybuty, takie jak wiek, płeć i ekspresję. Ma to zastosowanie wirtualnej rzeczywistości i tworzeniu postaci do gier.

Formalizacja rozwiązania, opis modelu lub adaptacji algorytmu:

Wykorzystano CGANs, które są preferowane, ponieważ pozwalają na integrację określonych cech (warunków) podczas generowania obrazu. Warunki te są dodawane jako wejście do generatora i dyskryminatora. W celu zachowania tożsamości, do projektu CGAN włączany jest mechanizm (np. funkcje straty zachowujące tożsamość), który ma za zadanie utrzymywać charakterystyczne elementy identyfikacyjne. W kontekście zastosowań omówiono także inne warianty, takie jak DCGAN, WGAN-GP, PGGAN, CycleGAN i StyleGAN.

Dane wejściowe (forma, struktura):

Model CGAN jest szkolony na zbiorze danych obrazów twarzy różnych osób (np. CelebA dataset). W CGAN wejściem są losowy szum oraz zmienne warunkujące (conditional variables), takie jak tożsamość, wiek, płeć lub atrybuty. Zbiór CelebA zawiera 202 599 zdjęć twarzy z 40 binarnymi atrybutami.

Przebieg algorytmu:

Generator CGAN jest warunkowany na atrybutach, co pozwala kontrolować i precyzyjnie dostosowywać generowane twarze. Szkolenie polega na wykorzystaniu zróżnicowanego zestawu danych z etykietowanymi parametrami, aby CGAN mógł uchwycić złożone korelacje między warunkami wejściowymi a cechami twarzy. Kluczowym elementem jest stosowanie strat lub ograniczeń związanych z tożsamością, aby model generował realistyczne twarze, które wiernie odzwierciedlają unikalne cechy danej osoby.

Rezultat (Zastosowanie Przemysłowe/Biznesowe):

Generowanie ludzkich twarzy ma szerokie zastosowanie w wirtualnej rzeczywistości, tworzeniu postaci do gier wideo i sztuce cyfrowej. Przyczynia się też do ulepszenia systemów rozpoznawania twarzy poprzez augmentację danych. CGANs ułatwiają edycję atrybutów twarzy w czasie rzeczywistym (np. zmiana wieku, płci). StyleGAN jest przydatny w generowaniu deepfake w czasie rzeczywistym, mającym wpływ na rozrywkę i treści online.

NOTATKA 3:

Conditional Activation GAN: Improved Auxiliary Classifier GAN

Notka bibliograficzna: Jeongik Cho, Kyoungro Yoon, "Conditional Activation GAN: Improved Auxiliary Classifier GAN", IEEE Access, 2020, DOI:10.1109/ACCESS.2020.3041480

Omówienie artykułu:

Poruszony problem inżynierski, cel badawczy:

Celem jest rozwiązanie problemów związanych z pomocniczym klasyfikatorem GAN - Auxiliary Classifier GAN (ACGAN), najczęściej używanym wariantem cGAN. Problemy te obejmują konieczność dostosowania hiperparametru dla stosunku straty adwersarialnej do klasyfikacyjnej, słabą konwergencję (brak sensownych gradientów) na wczesnym etapie treningu, oraz tendencję generatora do ignorowania warunków wejściowych przy zastosowaniu normalizacji wsadowej (Batch Normalization, BN) w dyskryminatorze.

Formalizacja rozwiązania, opis modelu lub adaptacji algorytmu:

Zaproponowano Conditional Activation GAN (CAGAN) jako zamiennik ACGAN. CAGAN jest traktowany jako integracja wielu niezależnych GAN-ów (po jednym na każdy warunek), które dzielą wszystkie warstwy ukryte. Stratą CAGAN jest suma strat poszczególnych GAN-ów. Dzięki temu CAGAN używa tylko jednej straty (conditional activation loss), eliminując potrzebę dostosowywania stosunku straty. Ponadto, w celu rozwiązania problemu BN w dyskryminatorze, wprowadzono technikę Mixed Batch Training, która utrzymuje stały stosunek danych rzeczywistych do generowanych w każdej partii.

Dane wejściowe (forma, struktura):

Dane są parami: dane rzeczywiste (x) i wektor warunku binarnego (cnd).

Generator otrzymuje wektor szumu i wektor warunku docelowego. W eksperymentach użyto:

- MNIST: Zestaw niebalansowany (5500 zer, 500 pozostałych cyfr).
- Celeb A: Obrazy 64x64, warunki: "Black or Brown hair", "Male", "Smiling".

Przebieg algorytmu:

CAGAN trenuje wszystkie warunki jednocześnie, generując znaczące gradienty nawet na wczesnym etapie treningu, dzięki zastosowaniu zaawansowanej straty adwersarialnej (np. LSGAN lub WGAN-GP). Strata w CAGAN jest obliczana poprzez iloczyn wewnętrzny (element-wise sum of products) z wektorem warunku cnd. Mixed Batch Training jest stopniowo wprowadzany: na początku proporcje danych rzeczywistych do generowanych są skrajne (np. 100:0 i 0:100), a następnie stopniowo zmieniają się w każdej epoce do docelowej proporcji (np. 50:50).

Rezultat (Zastosowanie Przemysłowe/Biznesowe):

CAGAN wykazuje lepszą wydajność pod względem szybkości treningu (performance increase per epoch) i mniejszej liczby hiperparametrów do dostosowania w porównaniu ze zmodyfikowanym ACGAN. Mixed Batch Training skutecznie zapobiega ignorowaniu wektora warunkowego przez generatora, co

ma miejsce w cGAN z normalizacją wsadową. Ulepszenie algorytmiczne CAGAN i Mixed Batch Training mają bezpośrednie przełożenie na efektywność generowania danych w kontrolowanych warunkach, co jest kluczowe dla wszelkich aplikacji inżynierskich i biznesowych wykorzystujących cGAN (np. generowanie realistycznych twarzy z określonymi atrybutami).

NOTATKA 4:

Facial Image Generation using fully trained Generative Adversarial Network

Notka bibliograficzna:

Ninad Kulkarni, Swati Jagtap, Hitesh Patil, Shreya Deshpande, "Facial Image Generation using fully trained Generative Adversarial Network", 8th International Conference on Computing, Communication, Control and Automation (ICCUBE) 2024, DOI: 10.1109/ICCUBE61740.2024.10775059

Omówienie artykułu:

Poruszony problem inżynierski, cel badawczy:

Istniejące metody generowania obrazów twarzy są często ograniczone do generowania ich jedynie z losowego szumu, bez możliwości syntezy na podstawie specyficznych cech. Celem badawczym jest opracowanie technik Generative Adversarial Networks (GANs) do generowania obrazów twarzy uwzględniających atrybuty (attribute-aware face images), które pasują do z góry określonych kryteriów.

Formalizacja rozwiązania, opis modelu lub adaptacji algorytmu:

W pracy porównano i wykorzystano Generative Adversarial Networks (GANs), w tym Conditional GAN (CGAN), Wasserstein GAN (WGAN) i Deep Convolutional GAN (DCGAN). Zastosowano architekturę z trzema ułożonymi generatorami (stacked generators) do produkcji obrazów twarzy w progresywnie wyższych rozdzielczościach: 64×64 , 128×128 i 256×256 . Dodatkowo, zaproponowano nową stratę dopasowania obraz-atrybut (image-attribute matching loss), aby poprawić korelację między wektorem wejściowym a generowanymi obrazami.

Dane wejściowe (forma, struktura):

Model jest szkolony i testowany na zbiorze danych CelebA Face Dataset. Zestaw ten zawiera 202 599 obrazów twarzy z 10 177 tożsamościami i jest opatrzony adnotacjami cech twarzy (np. atrybuty binarne). Dane wejściowe dla generatora to losowy szum oraz wektor cech przekonwertowany z tekstu wejściowego, co pozwala na generację kontrolowaną atrybutami.

Przebieg algorytmu:

Metoda zaczyna się od konwersji tekstu wejściowego na tablicę jego cech, które następnie są mapowane na GAN. W przypadku CGAN informacja warunkowa (y) jest integrowana zarówno z generatorem, jak i dyskryminatorem, co zapewnia kontrolę nad wyjściem. Następnie generatory (ułożone w stos) stopniowo zwiększają rozdzielczość. Modele są porównywane na podstawie kryteriów oceny wydajności, takich jak Inception Score, strata generatora i strata dyskryminatora.

Rezultat (Zastosowanie Przemysłowe/Biznesowe):

Generowane obrazy wysokiej jakości mogą służyć do augmentacji istniejących zbiorów danych obrazów, co redukuje potrzebę trenowania modeli na dużych zbiorach danych i zwiększa dokładność w praktycznych zastosowaniach, takich jak rozpoznawanie twarzy. Eksperymenty wykazały, że model CGAN osiąga lepsze wyniki w generowaniu obrazów twarzy w porównaniu z WGAN i DCGAN, zwłaszcza w zakresie straty generatora.