

Głęboki generator realistycznych krajobrazów i scen

dr inż. Paweł Kapusta

pawel.kapusta@p.lodz.pl

W ciągu ostatnich miesięcy nastąpił gwałtowny rozwój algorytmów pozwalających na generowanie obrazów i filmów z wykorzystaniem metod sztucznej inteligencji. Rozwiązania takie jak Stable Diffusion czy Midjourney pozwalają na tworzenie nowych treści, z wykorzystaniem jedynie opisów języka naturalnego. Algorytmy te mają jednak pewną wadę – nie dają nam pełnej kontroli nad generowanym obrazem.

Problem generowania syntetycznych krajobrazów czy całych scen jest poruszany w literaturze od kilku lat. Jednym z pierwszych algorytmów, dających ciekawe rezultaty, był algorytm GauGAN [1], opracowany przez zespół badaczy firmy Nvidia i wciąż rozwijany (GauGAN 2 [2] i GauGAN360 [3]). Umożliwił on generowanie krajobrazów na podstawie narysowanej wcześniej „maski”. Kolejnym krokiem było wykorzystanie dodatkowych algorytmów, takich jak CLIP, do analizy przestrzeni rozwiązań w taki sposób, aby uzyskać pewną kontrolę nad procesem generacji obrazów poprzez dodanie opisów tekstowych [5]. Poprawę jakości tych rozwiązań udało się uzyskać poprzez zastosowanie nowej grupy modeli sztucznej inteligencji – modeli dyfuzyjnych oraz zastosowanie mechanizmu uwagi (ang. *attention*) [6, 7].

Celem projektu jest przygotowanie rozwiązania, które pozwoli na generowanie scen i krajobrazów, przy zachowaniu jak największej kontroli nad tłem oraz obiektami znajdującymi się na pierwszym planie. Docelowo rozwiązanie powinno też pozwolić na modyfikację wygenerowanych wcześniej lub rzeczywistych obrazów poprzez np. dodawanie obiektów, modyfikację tła czy zmianę „stylu” obrazu. Tak przygotowany model może stanowić podstawę do zastosowania np. w grach komputerowych czy przemyśle filmowym.

Literatura:

1. Taesung Park, Ming-Yu Liu, Ting-Chun Wang, Jun-Yan Zhu (2019), Semantic Image Synthesis with Spatially-Adaptive Normalization
2. Nvidia Research, GauGAN 2 (2021), <http://imaginaire.cc/gaugan2/>
3. Nvidia Research, GauGAN 360 (2022), <http://imaginaire.cc/gaugan360/>
4. Federico Galatolo, Mario Cimino, Gigliola Vaglini (2021), Generating images from caption and vice versa via CLIP-Guided Generative Latent Space Search
5. Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, Mark Chen (2022), Hierarchical Text-Conditional Image Generation with CLIP Latents
6. Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, Daniel Cohen-Or (2022). PROMPT-TO-PROMPT IMAGE EDITING WITH CROSS-ATTENTION CONTROL
7. Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer (2021), High-Resolution Image Synthesis with Latent Diffusion Models