|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **2023 ISDJ 학교자율과정 프로젝트형 보고서** | | | | |
| 학 번 | 2학년 13반 20번 | 이 름 | 이재현 | |
| **보고서**  **제 목** | About the Mathematical Basis of Diffusion Models  (확산 모델의 수학적 기반에 대한 탐구) | **탐구분야**  (언어,수학,사회,과학,정보,예술,체육 中 택) | 수학, 과학, 정보 | |
| **목차**   1. 서론    1. 연구 동기 및 목적 2. 인공지능과 생성 모델링에 관한 역사    1. 뉴럴 네트워크(ANN)와 딥 뉴럴 네트워크(DNN)의 발전       1. The Introduction of Perceptron and the Pre-Deep Learning Era       2. 2010s - The Era of Deep Neural Networks       3. 2020s~Current – Advance to Large Scale Era    2. 생성 모델링의 발전       1. Deep Learning Era          1. Likelihood-based Models          2. Implicit Generative Models       2. Large Scale Era (2020s to Current)          1. Likelihood-based Models          2. Implicit Generative Models 3. 확산 모델에 대한 탐구    1. Prerequisites for understanding the Diffusion Model       1. Mathematical Intricacies          1. Linear Algebra             1. Basic Tensor Operations             2. Matrix Inverses, Eigenvalues and Eigenvectors             3. Matrix Factorization Techniques          2. Probability Theory and Statistics             1. Statistical distributions and Statistical Inference             2. Maximum Likelihood Estimation          3. Markov Chains and Markov Processes             1. Markov Chains             2. Markov Reversibility and Detailed Balance    2. The Evolution of Diffusion Model       1. 확산 모델의 필수 논문들          1. Denoising Diffusion Probabilistic Models          2. Diffusion Models Beat GANs on Image Synthesis          3. Denoising Diffusion Implicit Models       2. 모델 조건화와 잠재 확산 모델로의 전환          1. Adding Conditional Control to Text-to-Image Diffusion Models          2. High-Resolution Image Synthesis with Latent Diffusion Models       3. 이미지 외의 분야에서의 응용          1. Video Diffusion Models          2. DiffRF: Rendering-Guided 3D Radiance Field Diffusion          3. Noise2Music: Text-conditioned Music Generation with Diffusion Models 4. 확산 모델의 구현    1. Implementing the Denoising Diffusion Probabilistic Model (DDPM)       1. Transformer에서의 차용 – Sinusoidal Embeddings and Attention Module       2. Diffusion Process의 구현과 Training 5. 결론    1. 결론과 향후 탐구 방향 | | | |

|  |
| --- |
|  |
| **1. 서론**  **가. 연구 동기 및 목적** |
| **최근 인공지능 분야에서 갑자기 두각을 드러내어 세계 최대의 기술 기업들 – Alphabet Inc., Microsoft Corp., Meta Platforms Inc., Nvidia Corp. 등이 기술 개발과 투자를 진행하고 있는 하위 분야는 Generative Artificial Intelligence, 즉 생성형(生成形) 인공지능이다. 생성형 인공지능은 특정한 형식의 미디어 – 예를 들어, 텍스트, 이미지, 오디오, 동영상, 3D 모델 등을 인공신경망을 통해 양질로 생성하는 신경망들을 지칭한다.**  **이 하위 분야에서 이전 10년 동안 굉장한 많은 패러다임 변화와 그에 따른 발전 양상이 나타났다. 1982년 처음 고안되어 자연어 처리에 가장 많이 사용되던 Recurrent Neural Networks (RNN)** [1] **가족 (family)의 네트워크들** [2][3] **이 이들의 근본적인 문제들 – 병렬 연산 불가함과 기울기 소실을 해결한 Transformer** [4] **가 고안되어 대체되었고, 이미지 생성에 가장 많이 사용되던 Generative Adversarial Network (GAN)** [5] **가족의 네트워크들 역시 Diffusion Model (확산 모델)** [6] **의 등장과 함께 대체되는 추세다.**  **작년 나는 상대적으로 간단한 인공신경망들 – Residual Network (ResNET), U-Net, Autoencoder (AE, 오토인코더) 등을 구현하였는데, 이번 소논문에서는 발전된 주제를 세워, Diffusion Model의 수학적 기반, 모델 구조의 세부, 그리고 구현에 대해 연구해 볼 예정이다.** |

|  |
| --- |
| **나. 연구목적(필요성)** |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
| **2023 ISDJ 학교자율과정 프로젝트형 보고서** (2차\_본론) | | | |
| 학 번 | 2학년 13반 20번 | 이 름 | 이재현 |
| **보고서**  **제 목** | About the Mathematical Basis of Diffusion Models 확산 모델의 수학적 기반에 대한 탐구 | 지도교사 |  |
| **2. 본론**  다 레이아웃이에요 최종 아님  **2.1) The Amelioration of Neural Networks, Generative Modeling and the Establishment of Diffusion Models**  Lorem ipsum 어쩌구 저쩌구  2.1.1) Pre-Deep Learning Era:  - 1943: McCulloch-Pitts neuron model  - 1950: Alan Turing's "Computing Machinery and Intelligence"  - 1956: Dartmouth Workshop and the birth of AI as a field  - 1957: Introduction of the Perceptron  2.1.2) Deep Learning Era:  2.1. Rise of Deep Neural Networks:  - 2012: AlexNet wins ImageNet competition, popularizing deep convolutional neural networks (CNNs)  - 2014: Generative Adversarial Networks (GANs) introduced by Ian Goodfellow  - 2014: Google's DeepMind develops Deep Q-Network (DQN) for reinforcement learning  - 2015: LSTM networks demonstrate breakthroughs in sequence learning and natural language processing  - 2016: AlphaGo defeats world champion Go player, highlighting the power of deep reinforcement learning  2.2. Advancements in Deep Learning:  - 2017: Transformer model introduced, revolutionizing natural language processing (NLP)  - 2018: OpenAI introduces GPT (Generative Pre-trained Transformer) for language generation  - 2019: DeepMind's AlphaStar beats professional players in the game of StarCraft II  - 2019: StyleGAN generates highly realistic synthetic images  3. Large Scale Era (2020s to Current):  - 2020: GPT-3, OpenAI's large-scale language model with 175 billion parameters, demonstrates remarkable language generation capabilities  - 2020: Advances in computer vision with models like EfficientNet and ViT (Vision Transformer)  - 2020: DeepMind's AlphaFold makes significant progress in protein folding prediction  - 2021: Introduction of DALL-E, a model capable of generating images from textual descriptions  - Ongoing research and development of even larger and more efficient models, addressing ethical considerations and biases  It's important to note that this categorization is a simplified overview of the advancements in Artificial Intelligence, and there may be additional breakthroughs and events that could be included. The field of AI is rapidly evolving, and new developments continue to shape the landscape.  **2. 2.) The Development of Generative Modeling**  어쩌구 저쩌구 Dolor sit amet  2.2.1. Deep Learning Era:  2.2.1.1. Likelihood-based Models:  2013: Variational Auto-encoders (VAEs) proposed by Diederik P. Kingma and Max Welling  2014: Generative Moment Matching Networks (GMMN)  2015: Neural Autoregressive Distribution Estimator (NADE)  2015: PixelRNN and PixelCNN for generating images pixel by pixel  2.2.1.2 Implicit Generative Models:  2014: Generative Adversarial Networks (GANs) introduced by Ian Goodfellow  2016: Variational Inference GANs (VAEGAN)  2016: Adversarial Autoencoders (AAE)  2017: Boundary Equilibrium GANs (BEGAN)  2017: Wasserstein GANs (WGANs)  2017: Energy-Based GANs (EBGANs)  Large Scale Era (2020s to Current):  2.2.2.1. Likelihood-based Models:  2018: Autoregressive Transformers (e.g., GPT, GPT-2) for language generation  2020: VQ-VAE (Vector Quantized Variational Autoencoder)  2020: Flow-based generative models (e.g., Glow, FFJORD)  2.2.2.2. Implicit Generative Models:  2020: BigGAN and Deep Convolutional Inverse Graphics Network (DC-IGN)  2020: StyleGAN and StyleGAN2 for highly realistic image synthesis  2020: CLIP (Contrastive Language-Image Pretraining) for multimodal understanding  2020: Diffusion Models (e.g., DALL-E, Image GPT) for sequential generation and image synthesis | | | | |

|  |
| --- |
| **3.1. Prerequisites for understanding the Diffusion Model:**  3.1.1. Mathematical Intricacies:  3.1.1.1. aLinear Algebra:  3.1.1.1.1. Basic Tensor Operations:  - Understanding basic tensor operations, such as addition, multiplication, and transpose, is essential as tensors are fundamental data structures in many mathematical and machine learning operations.  3.1.1.1.2. Matrix Inverses, Eigenvalues, and Eigenvectors:  - Knowledge of matrix inverses, eigenvalues, and eigenvectors is important as they are often involved in solving equations and analyzing the properties of linear systems.  3.1.1.1.3. Matrix Factorization Techniques:  - Familiarity with matrix factorization techniques, such as singular value decomposition (SVD) and eigenvalue decomposition, is valuable as they play a crucial role in various mathematical operations and modeling approaches.  3.1.1.2. Probability Theory and Statistics:  3.1.1.2.1. Statistical Distributions and Statistical Inference:  - Understanding different statistical distributions (e.g., Gaussian, Poisson) and concepts of statistical inference (e.g., hypothesis testing, confidence intervals) is important as diffusion models often rely on probabilistic frameworks.  3.1.1.2.2. Maximum Likelihood Estimation:  - Knowledge of maximum likelihood estimation (MLE) is beneficial as it is a common technique for estimating model parameters and plays a key role in training and inference in diffusion models.  3.1.1.3. Markov Chains and Markov Processes:  3.1.1.3.1. Markov Chains:  - Understanding the concept of Markov chains, which are stochastic processes with the Markov property, is crucial as diffusion models utilize Markov chains as the underlying framework.  3.1.1.3.2. Markov Reversibility and Detailed Balance:  - Familiarity with the concepts of Markov chain reversibility and detailed balance is important as these properties are essential for understanding the behavior of diffusion models and the reversibility of the underlying Markov chains.  **3.2. The Evolution of Diffusion Models**  3.2.1. Essential Papers on Diffusion Models:  3.2.1.1. "Denoising Diffusion Probabilistic Models": This paper introduces denoising diffusion probabilistic models, which utilize the diffusion process to model probability distributions. It explores the denoising score matching framework for training these models effectively.  3.2.1.2. "Diffusion Models Beat GANs on Image Synthesis": This paper compares diffusion models with generative adversarial networks (GANs) for image synthesis tasks and demonstrates that diffusion models outperform GANs in terms of sample quality and training stability.  3.2.1.3. "Denoising Diffusion Implicit Models": This paper focuses on denoising diffusion implicit models, which learn probability distributions without explicitly modeling the diffusion process. It investigates the denoising score matching method for training these models.  3.2.2. Conditioning and Transitioning to Latent Diffusion Models:  3.2.2.1. "Adding Conditional Control to Text-to-Image Diffusion Models": This paper introduces the concept of conditional control in text-to-image diffusion models. It explores how conditioning the diffusion process on text inputs can generate images that align with specific textual descriptions.  3.2.2.2. "High-Resolution Image Synthesis with Latent Diffusion Models": This paper presents latent diffusion models for high-resolution image synthesis. It focuses on capturing and generating detailed and realistic images by incorporating the latent space diffusion process.  3.2.3. Applications in Fields Beyond Images:  3.2.3.1. "Video Diffusion Models": This paper extends diffusion models to the domain of video synthesis and generation. It explores the application of diffusion models to generate realistic videos by modeling the temporal dependencies of video frames.  3.2.3.2. "DiffRF: Rendering-Guided 3D Radiance Field Diffusion": This paper introduces diffusion models in the context of 3D radiance field synthesis. It explores how diffusion models can be guided by rendering techniques to generate realistic 3D scenes.  3.2.3.3. "Noise2Music: Text-conditioned Music Generation with Diffusion Models": This paper applies diffusion models to the generation of music conditioned on textual descriptions. It explores how diffusion models can generate music samples that align with specific textual prompts. |

|  |  |  |  |
| --- | --- | --- | --- |
| **2023 ISDJ 학교자율과정 프로젝트형 보고서** (3차\_본론 & 결론) | | | |
| 학 번 | 학년 반 번 | 이 름 |  |
| **보고서**  **제 목** |  | 지도교사 |  |
| 3.1.2.1. Essential Papers in Diffusion Models:  3.1.2.1.1. Denoising Diffusion Probabilistic Models:  This paper introduces the concept of denoising diffusion probabilistic models, which use diffusion processes to generate samples from a target distribution by denoising noisy samples.  3.1.2.1.2. Diffusion Models Beat GANs on Image Synthesis:  This paper explores the effectiveness of diffusion models compared to Generative Adversarial Networks (GANs) for image synthesis tasks, showcasing the advantages of diffusion models in generating high-quality images.  3.1.2.1.3. Denoising Diffusion Implicit Models:  This paper focuses on denoising diffusion implicit models, which are an extension of diffusion probabilistic models. It discusses techniques to improve the sampling quality and efficiency of diffusion models.  3.1.2.2. Transitioning to Conditional Modeling and Latent Diffusion Models:  3.1.2.2.1. Adding Conditional Control to Text-to-Image Diffusion Models:  This paper presents methods to incorporate conditional control into text-to-image diffusion models, enabling the generation of images conditioned on specific text descriptions.  3.1.2.2.2. High-Resolution Image Synthesis with Latent Diffusion Models:  This paper explores the use of latent diffusion models for high-resolution image synthesis. It discusses techniques for generating high-quality images at larger scales using diffusion models.  3.1.2.3. Applications Beyond Images:  3.1.2.3.1. Video Diffusion Models:  This paper extends diffusion models to the domain of video generation. It discusses techniques for generating realistic and coherent videos using diffusion processes.  3.1.2.3.2. DiffRF: Rendering-Guided 3D Radiance Field Diffusion:  This paper explores the application of diffusion models in 3D graphics and rendering. It introduces DiffRF, a method that leverages diffusion processes for rendering high-quality 3D scenes.  3.1.2.3.3. Noise2Music: Text-conditioned Music Generation with Diffusion Models:  This paper focuses on the application of diffusion models for text-conditioned music generation. It presents Noise2Music, a model that utilizes diffusion processes to generate music conditioned on textual prompts.    (※ 양식 변경 가능) | | | | |

|  |
| --- |
| **3. 결론** |
| **4. 참고문헌**  (※ 양식 변경 가능) |

# References

|  |  |
| --- | --- |
| [1] | John Hopfield, "Neural networks and physical systems with emergent collective computational abilities.," *Proceedings of the National Academy of Sciences of the United States of America,* 1982. |
| [2] | S. Hochreiter, "Long Short-Term Memory," *Neural Computation,* 1997. |