

작품 기술서

1. 작업 소개

작업 제목: $\Delta w \in \mathbb{R}^n$

작가: 김정훈 Kim Junghun

작업의 기본 개념과 출발점

$\Delta w \in \mathbb{R}^n$ 은 인공 신경망이 한 사람의 텍스트를 학습하는 과정을 시각화한 작품이다. Δw 는 가중치의 변화량(delta weight), 즉 학습의 최소 단위를 의미하며, \mathbb{R}^n 은 이 변화가 존재하는 수학적 공간의 차원을 뜻한다.

처음에는 록카쿠 아야코(Rokkaku Ayako) 스타일의 이미지를 AI로 생성하는 것으로 시작했다. 그러나 '록카쿠 스타일의 이미지'는 나오지만 거기에 록카쿠 자신은 없다는 것을 깨달았다. 표면만 닮았을 뿐, 그 사람이 왜 그렇게 그리는지에 대해서는 아무것도 담겨 있지 않았다.

록카쿠는 인터뷰에서 자신의 작업을 "그냥 같은 행위를 반복했을 뿐"이라고 말한다. AI의 학습도 마찬가지로 — 같은 연산을 반복하고, 어느 순간 무언가를 아는 상태에 도달한다. 둘 다 과정 자체에 의미가 있다. 이 작품은 그 과정을 기록한다.

전시 데이터와의 연결 방식

록카쿠 아야코가 직접 쓴 텍스트 9편(작업 과정에 대한 회상, 색에 대한 감각, 어린 시절 기억 등)을 doc-to-lora(SakanaAI) 기술에 입력하여 2,396,160개의 신경망 가중치로 변환했다. 이 숫자들이 록카쿠의 텍스트가 AI 내부에서 갖는 형태이다. 아무것도 모르는 빈 모델(0)이 이 목표값을 향해 학습하는 500단계의 과정을 세 가지 차원에서 기록하였다.

2. 작업 의도 및 해석 방향

주목한 요소

록카쿠의 작업에서 가장 주목한 것은 '반복'과 '직접성'이다. 록카쿠는 붓을 사용하지 않고 손으로 직접 그린다. 도구나 매개 사이에 아무것도 두지 않는다. AI 학습도 마찬가지로, 그래디언트 디센트(gradient descent)라는 동일한 연산을 수백 번 반복하며 가중치를 조정한다. 이 반복의 구조적 유사성에 주목했다.

전시를 해석하고 번역한 방식

록카쿠의 예술적 과정을 시의 학습 과정과 병치시켰다. 화면에 보이는 값들은 실제 신경망 파라미터 값이다. 초기에는 프랙탈이나 유체 시뮬레이션 등 다양한 시각적 표현을 시도했으나, 데이터의 구조와 변화 자체가 충분히 시각적이라는 판단 아래 이를 걷어냈다.

해당 형식을 선택한 이유

세 가지 차원(\mathbb{R} , \mathbb{R}^2 , \mathbb{R}^3)으로 동일한 학습 과정을 보여주는 형식을 선택했다:

- $\Delta w \in \mathbb{R}$ — 가중치를 십진수 수열로 화면에 직접 표시. 검정 배경 위의 흰색 숫자들이 학습 과정에서 실시간으로 변화한다. 26개 트랜스포머 레이어의 구조가 수직으로 나열되며, 좌측의 구조선이 모델 아키텍처를 암시한다.
- $\Delta w \in \mathbb{R}^2$ — 26개 트랜스포머 레이어 \times 8 랭크를 가중치 값을 32종의 색상과 형태로 변환하여 2차원 평면에 배치. 학습 후반부로 갈수록 색상이 분화하며, 모델 내부의 레이어별 특성 차이가 색의 분포로 드러난다.
- $\Delta w \in \mathbb{R}^3$ — 26개 레이어를 반투명 3D 슬래브로 공간에 배치하고 카메라가 궤도를 그리며 촬영. 레이어 간의 깊이 관계가 드러난다. 록카쿠의 언어에서 비롯된 가중치들이 공간 안에서 겹쳐질 때, 뜻하지 않게 록카쿠의 색채가 나타난다. 숫자 안에 감각의 잔상이 남아 있는 것인지, 혹은 우리가 그렇게 읽는 것인지는 알 수 없다.

시간 축에는 로그 스케일 압축을 적용하여, 초반의 격렬한 변화(실제로는 1-2초 만에 끝남)를 확대하고, 후반의 미세 조정은 압축했다. 음향은 26개 레이어의 그래디언트 크기를 펜타토닉 스케일로 변환하여 학습의 동역학을 청각적으로 전달한다.

세 차원의 영상은 순차적으로 이어진다. \mathbb{R}^3 까지 감상한 뒤 다시 \mathbb{R} 으로 돌아오면, 처음에는 숫자의 나열로만 보이던 화면 안에서 레이어의 구조와 학습의 흐름이 읽히기 시작할 수 있다. 같은 데이터를 다른 차원에서 바라본 경험이 시선을 바꾼다.

3. 제작 과정

1. **텍스트 수집 및 전처리:** 록카쿠 아야코가 직접 작성한 텍스트 9편을 수집, 작가의 목소리만 추출하여 정제
2. **LoRA 가중치 추출:** doc-to-lora(SakanaAI) 기술을 사용하여 텍스트를 Gemma-2-2b-it 모델의 LoRA 어댑터 가중치로 변환. 26개 레이어 \times 8 랭크의 A, B 행렬(2,396,160개 파라미터) 추출
3. **학습 시뮬레이션:** 빈 모델(모든 값이 0)에서 시작하여 추출된 목표 가중치를 향해 Adam 옵티마이저로 500단계 학습을 실행. 매 단계마다 전체 가중치 변화를 기록. 학습률 0.003, 그래디언트 노이즈 주입(scale 0.15, 학습 진행에 따라 감소)

4. 시각화 렌더링:

5. \mathbb{R} : Python + Pillow(DejaVu Sans Mono) 기반 커스텀 렌더러. 세로 캔버스에 네이티브 렌더링 후 가로/세로 비율별 크롭. 검정 배경, 흰색 숫자, 균일 밝기. 트랜스포머 아키텍처 구조선 포함
6. \mathbb{R}^2 : 32종 색상 + 32종 8×8 픽셀아트 마스크로 인코딩, 쿼터일 기반 매핑. 학습 과정에서의 가중치 분포 변화가 색상의 분화로 나타남
7. \mathbb{R}^3 : ModernGL(OpenGL) GPU 가속 렌더링, 26개 반투명 텍스처 쿼드, 카메라 궤도 회전, 블룸 후처리. $3840 \times 2160(4K)$, 60fps, HDR10
8. **음향 합성**: 26개 레이어의 그래디언트 크기를 펜타토닉 스케일 주파수로 변환, 사인파 합성(기본파 + 2차/3차 하모닉), 학습 진행에 따라 진폭 변조
9. **후반 작업**: 세 영상을 순차 연결, 페이드 투 블랙(1.5초) 트랜지션 적용, 엔딩 크레딧(나눔고딕코딩) 추가, HEVC HDR10 인코딩
10. **전시 설치 구성**: 프로젝터 1대(가로형, $\mathbb{R} \rightarrow \mathbb{R}^2 \rightarrow \mathbb{R}^3$ 순차 재생) + 세로형 모니터 3대(각각 \mathbb{R} , \mathbb{R}^2 , \mathbb{R}^3 독립 루프 재생). 세로 모니터에서는 각 차원의 학습 과정이 동기화 없이 독립적으로 반복되며, 관객은 세 화면을 동시에 조망할 수 있다.

4. 기술적 설명

사용 툴, 프로그램, 모델, 워크플로우

구분	도구
AI 모델	Gemma-2-2b-it (Google, 26억 파라미터 트랜스포머)
LoRA 변환	doc-to-lora (SakanaAI) — 텍스트 → 신경망 가중치
학습 프레임워크	PyTorch + CUDA (NVIDIA RTX 4090)
2D 렌더링	Python, NumPy, Pillow, OpenCV
3D 렌더링	ModernGL (OpenGL), 커스텀 GLSL 셰이더
음향 합성	SciPy, soundfile
영상 인코딩	FFmpeg (H.265/HEVC, HDR10 BT.2020/PQ)
색보정	FFmpeg curves/eq 필터
크레딧 폰트	나눔고딕코딩 (NanumGothicCoding)

이미지/영상 생성 방식

데이터 파이프라인:

록카쿠 텍스트 9편

- doc-to-lora (SakanaAI)
- LoRA 가중치 (26 layers × 8 ranks × 11,520 params = 2,396,160 values)
- Adam 옵티마이저 학습 시뮬레이션 (500 steps, LR=0.003)
- 프레임별 가중치 스냅샷 기록
- 시각화 렌더링 (\mathbb{R} / \mathbb{R}^2 / \mathbb{R}^3)
- 음향 합성
- 색보정 및 최종 영상 인코딩

시각 인코딩 체계 (\mathbb{R}^2 , \mathbb{R}^3): - 32종 8×8 픽셀아트 스프라이트 (다이아몬드, 십자, 하트, 별 등) - 32종 네온 컬러 (빨강 → 주황 → 노랑 → 초록 → 시안 → 파랑 → 보라 → 핑크 스펙트럼) - 가중치 값 → 퀀타일 기반 31개 경계로 32개 구간 분류 → 스프라이트 종류 결정 - 가중치 크기(절대값) → 밝기(brightness = $0.7 + 0.3 \times$ 정규화된 크기)

3D 렌더링 파라미터 (\mathbb{R}^3): - 26개 레이어: 간격 0.6, 너비 4.0, 높이 3.0 - 카메라: 거리 12.0, 시각 각도 1.2 rad, 궤도 속도 0.35, 높이 진동 ± 2.0 - 레이어당 3,072개 스프라이트 (64×48 그리드), 총 79,872개 - 블룸: 추출 임계값 0.4, 합성 강도 0.6, 1/4 해상도 가우시안 블러 - 알파 블렌딩: 거리 기반 투명도, 뒤에서 앞 순서 렌더링

편집 및 후반 작업 과정

- **영상 통합:** 세 작품(\mathbb{R} , \mathbb{R}^2 , \mathbb{R}^3)을 하나의 타임라인으로 순차 연결, 페이드 투 블랙(1.5초) 트랜지션
- **\mathbb{R}^2 색보정:** FFmpeg curves 필터를 활용한 색상 톤 보정
- **비율 변환:** 세로(9:16) 영상 → 가로(16:9) 프레임에 맞춰 확대, 상하 크롭
- **색공간 변환:** HDR10(BT.2020/PQ) ↔ SDR(BT.709) 양방향 변환
- **엔딩 크레딧:** 검정 배경, 중앙 정렬 흰색 텍스트, 나눔고딕코딩 폰트
- **최종 인코딩:** HEVC(H.265) 10-bit, HDR10(BT.2020/PQ), CRF 14-16