AI Seminar Report

Date: 03/28/2025

Name	Vo Hoang Chuong (우황장)	**Student No.**	2024121193
School	Dongguk University	**Course Classification**	Ph.D.
Speaker Name	진경환 부교수	**Speaker Affiliation**	고려대학교
Title	Towards Lossless Digital Imaging World: Quantization-Aware Deep Neural Networks		

■ Seminar Contents

I. Foundational Concepts: Continuous vs. Discrete Representation

- **Representation Domains:** We distinguish between the continuous domain, reflecting real-world signals, and the discrete domain, representing digital information.
- **Domain Transition:** Mapping from the continuous to the discrete typically involves sampling followed by quantization.
- **Quantization:** Formally, quantization is a function mapping a continuous input space (X) to a finite set of discrete output levels (Q_n), often utilizing methods like ceiling quantization to discretize analog functions.
- **Quantization-Aware Processing:** Systems designed for the discrete world often incorporate both quantization and de-quantization steps, particularly relevant for numerical representations like FP16/FP32 floating-point formats.

II. Paper 1: Lossless Implicit Neural Representations via Bit Plane Decomposition (CVPR 2025)

- **Objective:** To achieve lossless signal representation using Implicit Neural Representations (INRs).
- **Methodology:** Proposes Bit-Plane Decomposition, where an INR maps input coordinates (and potentially associated bit coordinates) to individual bit planes of the target signal.
- **Training Paradigm:** The INR is supervised to reconstruct the original signal accurately from the predicted bit planes.
- **Theoretical Contributions:**
    - Establishes an upper bound on the necessary INR parameter count relative to the target bit precision.
    - Identifies and analyzes "bit bias" and "bit-spectral bias," phenomena within the bit-plane domain analogous to spectral bias in standard INRs.

- **Key Findings & Performance:**
  - Achieves perfect reconstruction (infinite PSNR) for low bit depths (1-4 bits) and exceptionally high fidelity (PSNR > 100) for 8 bits.
  - Observes faster training convergence for lower bit-precision targets.
  - Demonstrates superior performance compared to state-of-the-art methods in terms of convergence speed and reconstruction accuracy.
- **Significance & Applications:** The approach facilitates lossless representation in digital media, with potential applications in ternary networks, lossless compression algorithms, and bit-depth expansion tasks.

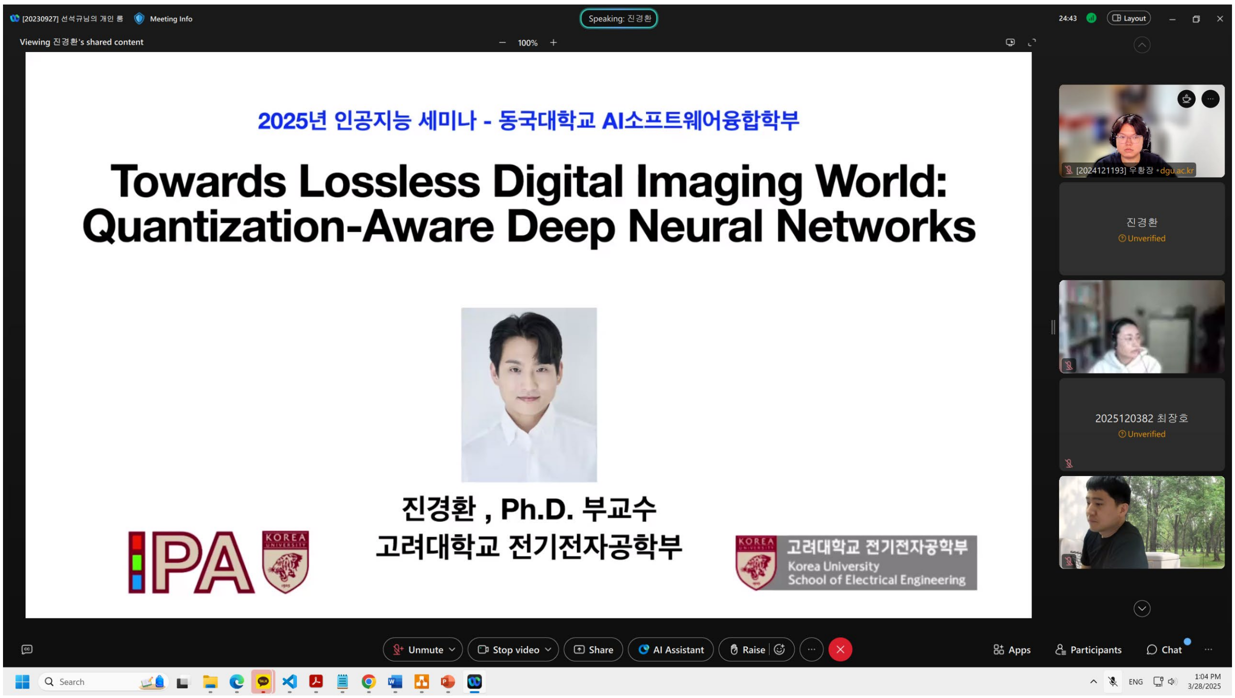
III. Paper 2: Grid-wise Addressable Memory (GAM) (submitted)

- **Objective:** To develop 'Memory-Integrated' Implicit Neural Representations that are capable of lossless data storage and retrieval.
- **Conceptual Framework:** Integrates memory directly within the INR, visualized as storing data points within a stable energy landscape.
- **Methodology & Theory:**
  - Employs principles derived from Lyapunov stability theory to ensure reliable data retrieval.
  - Utilizes recurrent stable training dynamics, potentially enforcing monotonic gradient descent for convergence to stored states.
  - Investigates quantized representations within GAM, analyzing the interplay between bit-level discontinuities and smoother gray-level representations.
- **Key Findings & Performance:**
  - Realizes stable, addressable memory functionality within the implicit function.
  - Achieves perfect reconstruction of stored data.
  - Demonstrate training efficiency, requiring fewer iterations to attain high PSNR compared to alternative approaches.
- **Applications:**
  - Successfully applied to representing complex signals, such as Signed Distance Functions (SDFs).
  - Shown to be adequate for tasks such as Novel View Synthesis, leveraging its data retrieval capabilities.

■ What have you learned from this seminar?

- **Bridging Worlds:** I gained a clearer understanding of how we translate continuous, real-world information into the discrete, digital domain, primarily through sampling and quantization.
- **Lossless INRs:** Explored cutting-edge techniques using Implicit Neural Representations (INRs) to achieve lossless data representation, moving beyond typical approximation methods.
- **Bit-Plane Strategy:** Learned about a specific approach (Bit-Plane Decomposition) that cleverly breaks down signals into individual bit layers for an INR to learn, enabling perfect reconstruction, especially at lower bit depths.
- **Memory in INRs:** Understood a novel concept (Grid-wise Addressable Memory - GAM) that integrates memory directly within an INR, using stability principles for perfect data storage and retrieval.
- **Theoretical Insights:** Appreciated the theoretical work presented, including the establishment of bounds on model size based on precision and the identification of phenomena such as "bit bias."
- **Efficiency Matters:** It is noticeable that both presented methods emphasize efficient training, achieving high performance (such as perfect reconstruction or high PSNR) faster than previous approaches.
- **Practical Potential:** Recognized the broad applicability of these advanced INR techniques in areas such as lossless compression, complex signal representation (e.g., SDFs), and even novel view synthesis.

Capture your Webex screen showing the start time (1:00 PM).



Capture your Webex screen showing the end time as well.