学　位　論　文　内　容　の　要　旨

博士の専攻分野の名称　　博士（情報科学）　　氏名　蘭 子文

学　位　論　文　題　名

A study on latent representation learning for understanding, generation, and copyright protection of artificially created works

（人工創作物の理解・生成および著作権保護のための潜在表現学習に関する研究）

In recent years, generative artificial intelligence has made significant advances and has demonstrated remarkable capabilities across a broad spectrum of creative domains, including illustrations, graphic design, and various forms of artificially created works. Contemporary multimodal AI models are now capable of learning rich semantic representations and producing highly detailed and stylistically coherent outputs. These developments have led to the rapid spread of artificially created works and the integration of AI into professional and amateur creative workflows. However, as these models become increasingly prevalent in creative and practical workflows, three core challenges arise: (1) improving semantic understanding, (2) enhancing the flexibility and robustness of generative models, and (3) ensuring the safety and protection of creative content. These challenges are interconnected; systems that understand semantics more accurately can support more reliable generation, while both capabilities must be grounded in mechanisms that ensure safe and ethical use of creative data. Addressing these three challenges is essential for constructing a comprehensive and trustworthy framework for next-generation creative AI systems.

To tackle these issues, this thesis investigates latent representation learning as a unified foundation for improving semantic modeling performance, achieving controllable and robust generation, and preventing unauthorized use of creative content. Latent space plays a central role in modern generative models: it is where semantic concepts are formed, transformed, disentangled, and recombined. By focusing on this intermediate space rather than surface-level pixel information, the proposed framework provides a consistent and interpretable basis for addressing the three major challenges of creative AI.

The first challenge concerns semantic understanding performance. Artificially created works often contain stylized abstractions, symbolic relationships, hierarchical attributes, and strongly structured semantics that challenge traditional image recognition pipelines. As a result, traditional methods frequently struggle to interpret multi-label relationships or hierarchical semantics inherent in such content. Latent representation learning provides a promising direction by enabling models to capture high-level semantic structures and contextual dependencies that are not easily represented by pixel-level features.

The second challenge relates to generation flexibility and robustness. In practical use cases, especially when users repeatedly personalize models with new concepts, text-to-image diffusion models often encounter concept bleed-through, where newly added concepts interfere with previously learned ones. To provide a smoother narrative connection, this thesis focuses on scenarios in which users continually incorporate personalized concepts over time, such as creating multiple unique styles or expanding a

collection of custom entities. Existing continual learning approaches attempt to adapt model parameters but often fail to maintain clearly separated semantic representations in latent space. To address this issue, this thesis proposes a data-level adversarial perturbation method that shifts latent semantic representations during training, ensuring separability between old and new concepts without modifying model architecture.

The third challenge examines content safety and copyright protection. Conventional protection techniques, such as watermarking or pixel-level perturbations, do not effectively intervene in the latent learning mechanisms of diffusion models. To overcome this limitation, the thesis introduces a latent-space–oriented adversarial protection method that manipulates latent distributions in training images in a visually imperceptible manner. This disrupts the semantic extraction process and significantly reduces the ability of diffusion models to learn or reproduce protected styles or attributes, providing a practical defense mechanism for creators.

The structure of the thesis is outlined as follows. Chapter 1 describes the research background and motivation, emphasizing the need for latent-representation–based methods in understanding, generating, and protecting artificially created works. Chapter 2 summarizes related works on semantic understanding, generative modeling, and AI-based copyright protection, identifying the shortcomings of existing approaches. Chapter 3 presents a latent representation learning method for semantic understanding, focusing on semantic attribute extraction, hierarchical relationship modeling, and multimodal fusion. Chapter 4 introduces a latent semantic–based generation control method designed to prevent concept bleed-through and stabilize generative behavior in diffusion models. Chapter 5 proposes an adversarial-perturbation-based approach for copyright protection, preventing generative models from learning unauthorized artistic concepts. Chapter 6 concludes the thesis and discusses future research directions for latent representation learning in creative AI.

In summary, this thesis presents a unified framework that leverages latent representation learning to address the multifaceted challenges of understanding, generating, and protecting artificially created works. Through extensive evaluations across multi-label recognition, diffusion-based generation, and adversarial protection tasks, the proposed framework demonstrates strong improvements over existing approaches. By modeling creative content in latent space, the proposed methods provide deeper semantic insight, more reliable generative control, and enhanced protective capabilities against unauthorized usage.

This dissertation contributes to the foundation of responsible creative AI by proposing methods that respect both technological advancement and artistic integrity. The work emphasizes that latent representation learning not only enhances the performance of AI systems but also plays a crucial role in shaping an ethical and sustainable ecosystem for the future of artificially created content. As generative AI continues to expand into new domains, the unified framework presented in this thesis offers valuable direction for designing systems that balance expressive power, semantic coherence, and creator rights, thereby supporting the development of creative technologies that are both innovative and respectful of human artistry.