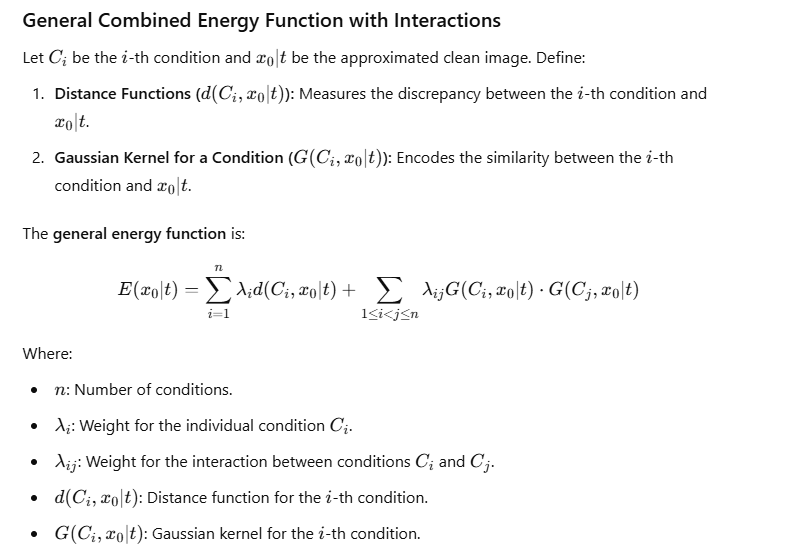
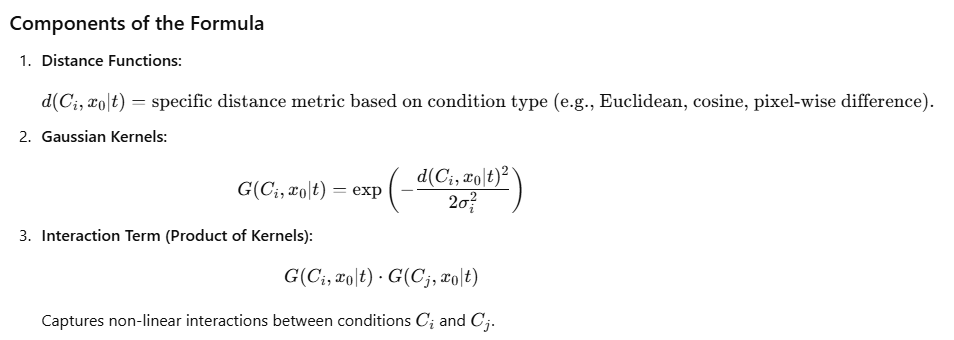
****

****

**Thesis Title:** Enhancing Multi-Conditional Training-Free Image Generation with Interaction-Aware Gaussian Kernels

**Abstract:** Multi-conditional image generation aims to synthesize images that satisfy diverse conditions such as textual descriptions, segmentation masks, and landmark constraints. Traditional approaches that use weighted sums of distance functions to approximate the energy function often produce suboptimal results, as they fail to capture complex, non-linear interactions between conditions. This thesis introduces a novel framework that incorporates interaction terms through the product of Gaussian kernels, enabling the modeling of interdependencies between conditions. The proposed method demonstrates superior performance in producing coherent and condition-consistent images compared to existing techniques.

**Chapter 1: Introduction**

1.1 **Problem Statement** Multi-conditional image generation has emerged as a critical area in computer vision, where the goal is to generate images that simultaneously satisfy multiple user-defined constraints. For example, generating a facial image that aligns with textual descriptions, adheres to geometric landmark constraints, and respects segmentation masks. Traditional approaches to this problem rely on approximating the energy function as a weighted sum of individual distance functions between the generated image and each condition. While straightforward, this method often leads to subpar results due to its inability to account for interactions between conditions. The absence of interaction modeling results in:

* Conflicting conditions being treated in isolation.
* Loss of coherence in generated outputs.
* Limited adaptability to complex, real-world multi-modal scenarios.

1.2 **Motivation** Real-world conditions often interact in non-linear and complex ways. For instance, facial landmarks inherently influence segmentation masks, and textual descriptions may dictate geometric features. Ignoring these interactions leads to incomplete modeling of the underlying problem. To address this, we propose incorporating interaction terms through Gaussian kernels, which provide a smooth, differentiable, and interpretable mechanism for capturing interdependencies between conditions.

1.3 **Contributions** This thesis makes the following contributions:

* Proposes a novel energy function combining weighted distance functions and Gaussian kernel-based interaction terms.
* Demonstrates the efficacy of the approach through applications such as face generation, sketch-based synthesis, and medical image reconstruction.
* Provides a modular framework extensible to new conditions and datasets without additional training.

**Chapter 2: Related Work**

2.1 **Traditional Multi-Conditional Image Generation** Discuss the weighted sum of distance functions approach and its limitations:

* Inability to model interactions.
* Sensitivity to conflicting conditions.
* Limited scalability to diverse tasks.

2.2 **Gaussian Kernels in Image Synthesis** Overview of Gaussian kernels and their application in similarity metrics and energy functions.

2.3 **Interaction Modeling in Vision Tasks** Explore methods that model condition interdependencies in other domains and their relevance to image generation.

**Chapter 3: Methodology**

3.1 **Proposed Energy Function** We define the energy function as:

Where:

* : Distance function for condition .
* : Gaussian kernel for condition .
* : Weight for individual conditions.
* : Weight for interactions between conditions.

3.2 **Modeling Distance Functions** Define condition-specific distance functions for:

* Text-image similarity (CLIP).
* Structural constraints (segmentation masks).
* Geometric alignment (landmarks).

3.3 **Interaction Terms with Gaussian Kernels** Illustrate the role of interaction terms and their ability to:

* Resolve conflicting conditions.
* Enhance coherence by leveraging interdependencies.

**Chapter 4: Experiments**

4.1 **Experimental Setup**

* Dataset descriptions (e.g., CelebA-HQ for facial synthesis).
* Baseline models for comparison.

4.2 **Evaluation Metrics**

* Perceptual quality (e.g., FID scores).
* Condition consistency (e.g., IoU for segmentation, alignment scores for landmarks).
* User studies for subjective evaluation.

4.3 **Results and Analysis**

* Comparison of traditional weighted sum and proposed Gaussian kernel methods.
* Ablation studies to demonstrate the impact of interaction terms.
* Visualization of generated outputs.

**Chapter 5: Discussion**

5.1 **Advantages of Interaction Modeling**

* Improved coherence and condition satisfaction.
* Flexibility to adapt to new conditions and tasks.

5.2 **Challenges and Limitations**

* Computational overhead due to interaction terms.
* Sensitivity to hyperparameters (, ).

**Chapter 6: Conclusion and Future Work**

6.1 **Conclusion** This thesis demonstrates that incorporating interaction terms through Gaussian kernels addresses the limitations of traditional weighted-sum approaches for multi-conditional image generation. The proposed method significantly enhances the quality and coherence of generated images by modeling interdependencies between conditions.

6.2 **Future Work**

* Explore higher-order interactions.
* Investigate adaptive weight learning for and .
* Extend the framework to other generative tasks such as video synthesis and 3D reconstruction.