DACS: Diffusion-based Aerobatics with CBF-Guided Sampling for Urban Air Vehicles
Comprehensive Analysis and Improvement Framework

AI Model Analysis

October 14, 2025

Abstract

This document provides a comprehensive analysis of DACS (Diffusion-based Aerobatics with CBF-Guided Sampling), an enhanced transformer-based diffusion model for generating aerobatic UAV trajectories in urban environments. The model incorporates obstacle awareness through MLP-based obstacle encoding, Control Barrier Function (CBF) guidance for safety guarantees, and multi-modal conditional inputs including target waypoints, maneuver styles, historical observations, and obstacle information. We examine the enhanced architecture, CBF-guided sampling methodology, performance characteristics, and propose improvement strategies for urban air mobility applications.

Contents

1 Introduction

The DACS model represents a significant advancement in UAV trajectory generation for urban environments, combining diffusion processes with transformer-based architecture and Control Barrier Function guidance. The model is designed to handle complex aerobatic maneuvers while maintaining physical constraints, obstacle avoidance, and maneuver style consistency in cluttered urban settings.

2 Model Architecture Analysis

2.1 Overall Framework

The DACS framework consists of four main components:

- **Obstacle-Aware Diffusion Transformer**: Core generative model with obstacle encoding

- **CBF-Guided Diffusion Process**: Safety-guaranteed sampling with barrier functions
- **Multi-modal Conditioning System**: Integrated obstacle, target, and style processing
- **Obstacle Encoder MLP**: Neural network for obstacle representation learning

2.2 Mathematical Formulation

2.2.1 Diffusion Process with CBF Guidance

The forward diffusion process follows the standard formulation:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1 - \bar{\alpha}_t)I) \tag{1}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ and $\alpha_t = 1 - \beta_t$.

The reverse process is enhanced with CBF guidance:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t) - \gamma_t \nabla V(x_t), \sigma_t^2 I) \tag{2}$$

where $V(x_t)$ is the control barrier function and γ_t is the guidance strength.

2.2.2 Control Barrier Function

The CBF for multiple spherical obstacles:

$$V(x) = \sum_{\tau=1}^{T} \sum_{i=1}^{N_{obs}} \max(0, r_i - \|p_\tau - c_i\|)^2 \tag{3}$$

where p_τ is the position at time τ, c_i and r_i are obstacle centers and radii.

2.2.3 Enhanced Transformer Architecture

The diffusion transformer employs obstacle-aware conditioning:

$$h_0 = \text{InputProj}(x) + \text{PosEnc} + \text{EnhancedCondEmbed}(t, target, action, obstacles) \tag{4}$$

$$h_l = \text{TransformerLayer}(h_{l-1}, h_{l-1}), \quad l = 1, \ldots, L \tag{5}$$

$$\hat{x}_0 = \text{OutputProj}(h_L) \tag{6}$$

2.3 Component Details

2.3.1 Obstacle Encoder MLP

The obstacle encoder processes multiple obstacles:

$$e_{obs} = \text{GlobalObstacleEncoder}\left(\text{AttentionAggregation}\left(\text{ObstacleMLP}(o_i)\right)_{i=1}^{N_{obs}}\right) \tag{7}$$

2.3.2 Enhanced Condition Embedding

Multi-modal conditioning with obstacle fusion:

$$e_{cond} = \text{FusionLayer}([e_t; e_{target}; e_{action}; e_{obstacles}]) \tag{8}$$

3 Model Configuration

Table 1: Enhanced Model Configuration Parameters

Parameter	**Value**	**Description**
Latent Dimension	256	Transformer hidden size
Number of Layers	4	Transformer decoder layers
Number of Heads	4	Multi-head attention heads
Dropout Rate	0.1	Regularization parameter
Diffusion Steps	30	Number of diffusion steps
Sequence Length	60	Trajectory time steps
State Dimension	10	$[speed, x, y, z, attitude(6)]$
History Length	5	Historical observations
Target Dimension	3	Waypoint coordinates
Action Dimension	5	Maneuver styles
Max Obstacles	10	Maximum number of obstacles
Obstacle Feature Dim	4	$[x, y, z, radius]$
CBF Guidance Gamma	100.0	Barrier guidance strength
Obstacle Radius	5.0	Safe distance radius

4 Strengths and Innovations

4.1 Architectural Advantages

1. **Obstacle-Aware Generation**: MLP-based obstacle encoding integrated into transformer
2. **CBF Safety Guarantees**: Formal safety verification through barrier functions
3. **Multi-scale Conditioning**: Effective integration of temporal, spatial, behavioral, and obstacle constraints
4. **Causal Attention**: Proper temporal modeling with masked self-attention
5. **Unified Loss Function**: Comprehensive loss with obstacle avoidance and continuity terms

4.2 Training Methodology

- **Enhanced Data Generation**: Diverse aerobatic maneuver simulation with urban obstacles

- **Progressive Normalization**: Dimension-aware normalization strategy
- **CBF-Guided Sampling**: Safety-aware reverse diffusion process
- **Obstacle-Aware Loss**: Training with explicit obstacle avoidance objectives

5 Limitations and Improvement Areas

5.1 Architectural Limitations

5.1.1 Obstacle Representation

- **Issue**: Fixed maximum number of obstacles (10)
- **Impact**: Cannot handle arbitrarily large obstacle sets
- **Solution**: Dynamic obstacle processing or graph neural networks

5.1.2 Attention Mechanism

- **Issue**: Full self-attention has $O(N^2)$ complexity
- **Impact**: Limits sequence length and training efficiency
- **Solution**: Implement sparse attention or linear attention variants

5.2 Training Limitations

5.2.1 Loss Function Complexity

- **Issue**: Multiple loss components require careful balancing
- **Impact**: Sensitive to hyperparameter tuning
- **Solution**: Adaptive loss weighting or multi-task learning

5.2.2 CBF Guidance Stability

- **Issue**: Fixed guidance strength γ_t
- **Impact**: May cause training instability or over-constraint
- **Solution**: Learnable guidance scheduling or adaptive γ_t

Algorithm 1 Dynamic Obstacle Encoding

```
procedure DynamicObstacleEncode(obstacles, k)
    embeddings ← ObstacleMLP(obstacles)
    indices ← TopKByDistance(embeddings, k)
    selected ← embeddings[indices]
    aggregated ← AttentionAggregate(selected) return aggregated
end procedure
```

Algorithm 2 Hierarchical Transformer Architecture

```
procedure HierarchicalEncode(x, levels)
    patches ← Patchify(x, patch_size)
    h ← LocalAttention(patches)
    for i ← 1 to levels do
        h ← GlobalAttention(h)
        h ← Downsample(h)
    end forreturn h
end procedure
```

6 Proposed Improvements

6.1 Architectural Enhancements

6.1.1 Dynamic Obstacle Processing

6.1.2 Hierarchical Transformer

6.2 Training Improvements

6.2.1 Adaptive CBF Guidance

Learnable guidance scheduling:

$$\gamma_t = \gamma_{base} \cdot \sigma(W_t t + b_t) \tag{9}$$

6.2.2 Enhanced Loss Balancing

Uncertainty-weighted multi-task loss:

$$\mathcal{L}_{total} = \sum_{i=1}^{D} \frac{1}{2\sigma_i^2} \mathcal{L}_i + \log \sigma_i \tag{10}$$

where σ_i are learnable uncertainty parameters.

6.2.3 Curriculum Learning

Progressive difficulty scheduling:

1. Phase 1: Basic trajectory learning without obstacles
2. Phase 2: Simple obstacle avoidance
3. Phase 3: Complex urban environments with multiple obstacles

6.3 Physical Constraints Integration

6.3.1 Enhanced Dynamics Constraints

$$\mathcal{L}_{physics} = \lambda_{vel}\|\hat{v} - v_{gt}\|^2 + \lambda_{acc}\|\hat{a} - a_{gt}\|^2 + \lambda_{jerk}\|\hat{j} - j_{gt}\|^2 \tag{11}$$

6.3.2 Urban Environment Constraints

$$\mathcal{L}_{urban} = \lambda_{building} \sum \max(0, h_{min} - z_t) + \max(0, z_t - h_{max}) \tag{12}$$

7 Experimental Evaluation Framework

7.1 Evaluation Metrics

Table 2: Enhanced Evaluation Metrics for Urban Scenarios

Metric	**Description**
ADE	Average Displacement Error
FDE	Final Displacement Error
Z-Axis MAE	Mean Absolute Error in Z-axis
Maneuver Fidelity	Style classification accuracy
Obstacle Clearance	Minimum distance to obstacles
Collision Rate	Percentage of colliding trajectories
Physical Plausibility	Dynamics constraint satisfaction
Diversity	Multi-modal distribution coverage
Success Rate	Obstacle avoidance success percentage

7.2 Ablation Studies

Recommended ablation studies for DACS:

1. Obstacle encoding ablation (MLP vs. simple concatenation)
2. CBF guidance ablation (with vs. without safety guidance)
3. Conditioning ablation (remove target/action/history/obstacles)
4. Architecture variants (different attention mechanisms)
5. Loss function components (obstacle term contribution)

8 Implementation Recommendations

8.1 Code Improvements

- **Modularization**: Separate obstacle processing, CBF guidance, and transformer components
- **Configuration Management**: Use config classes for urban scenario parameters

- **Visualization Tools**: Enhanced plotting for obstacle-aware trajectories
- **Logging**: Comprehensive training monitoring with safety metrics

8.2 Computational Optimization

- **Mixed Precision**: FP16 training for memory efficiency
- **Gradient Checkpointing**: Memory-efficient backpropagation
- **Distributed Training**: Multi-GPU support for large-scale urban scenarios
- **Obstacle Caching**: Efficient obstacle representation reuse

9 Urban Air Mobility Applications

9.1 Real-world Deployment Considerations

- **Sensor Integration**: Fusion with LiDAR, camera, and radar data
- **Real-time Performance**: Optimization for onboard computation
- **Regulatory Compliance**: Adherence to urban air traffic management
- **Uncertainty Handling**: Robustness to sensor noise and dynamic obstacles

9.2 Scalability to Complex Urban Environments

- **Multi-building Scenarios**: Handling urban canyons and complex geometries
- **Dynamic Obstacles**: Adaptation to moving vehicles and pedestrians
- **Weather Conditions**: Robustness to wind, precipitation, and visibility
- **Communication Constraints**: Operation in GPS-denied environments

10 Conclusion and Future Directions

The DACS model presents a comprehensive framework for urban aerobatic trajectory generation with several innovative features. The key strengths include obstacle-aware transformer architecture, CBF-guided safety guarantees, and unified training methodology. The main areas for improvement involve scalability to complex urban environments, real-time performance, and enhanced robustness.

Future research directions should explore:

- **Real-world Urban Deployment**: Transfer learning to actual urban environments
- **Online Adaptation**: Real-time obstacle avoidance and replanning
- **Multi-agent Coordination**: Swarm behavior in urban airspace

- **Advanced Obstacle Representations**: Signed distance fields and occupancy grids
- **Uncertainty Quantification**: Probabilistic safety guarantees
- **Human-in-the-Loop**: Interactive trajectory refinement

Appendix

A. Mathematical Derivations

A.1 CBF-Guided Reverse Diffusion

The CBF-guided reverse process derivation:

$$p_\theta(x_{t-1}|x_t) \propto p_\theta(x_{t-1}|x_t) \exp(-\gamma_t V(x_t)) \tag{13}$$

A.2 Obstacle Distance Gradient

Gradient of the barrier function:

$$\nabla V(x_t) = \sum_{i=1}^{N_{obs}} -2 \cdot \max(0, r_i - \|p_t - c_i\|) \cdot \frac{p_t - c_i}{\|p_t - c_i\|} \tag{14}$$

B. Hyperparameter Search Space

Enhanced hyperparameter ranges for urban scenarios:

- Learning rate: $[1e-5, 1e-3]$ (log scale)
- Latent dimension: $\{128, 256, 512\}$
- Number of layers: $\{4, 6, 8\}$
- Number of heads: $\{4, 8, 16\}$
- Diffusion steps: $\{50, 100, 200\}$
- CBF guidance strength: $[10.0, 500.0]$
- Obstacle weight: $[1.0, 20.0]$
- Maximum obstacles: $\{5, 10, 20, 50\}$

C. Urban Scenario Specifications

Table 3: Urban Environment Parameters

Parameter	Typical Range
Building Height	20-200 m
Street Width	15-30 m
Obstacle Density	5-50 obstacles/km^2
Minimum Clearance	5-20 m
Maximum Climb Rate	10 m/s
Urban Canyon Aspect Ratio	1:1 to 1:3