



**University of
Nottingham**

UK | CHINA | MALAYSIA

**Project Proposal:
Investigation of Defense Mechanisms against Model
Inversion Attacks**

Submitted **October, 2025**, in partial fulfillment of
the conditions for the award of the degree **BSc Computer Science**.

Yixuan ZHANG

20513731

hnyyz39@nottingham.edu.cn

Supervised by Dr. Jianfeng REN

BSc (Hons) Computer Science

School of Computer Science
University of Nottingham Ningbo China

Abstract

Contents

1	Introduction	4
1.1	Background and Motivation	4
1.2	Problem Statement	4
1.3	Aim and Objectives	4
2	Related Work	4
2.1	Model Inversion Attacks	4
2.2	Split Learning / Edge-Cloud Privacy	4
2.3	Conditional Entropy Minimization and Surrogates	4
2.4	Gated Attention Mechanisms	4
2.5	Slot Attention and Cross-Attention	4
3	Methodology	4
3.1	Baseline Overview (CVPR Spotlight)	4
3.2	Proposed Method 1: Gated-Attention CEM	4
3.3	Proposed Method 2: Slot + Gated Cross-Attention CEM (Exploratory)	4
3.4	Evaluation Protocol	4
4	Implementation	4
5	Preliminary Results and Analysis	4
5.1	Main Quantitative Comparison	4
5.2	Why Method 1 Improves Over the Baseline	4
5.3	Why Method 2 Underperforms So Far	4
6	Progress Against Workplan	4
6.1	Completed Work	4
6.2	Original Plan vs Current Status	5
6.3	Updated Plan	5
7	Reflection and Risk Management	5
7.1	Key Challenges So Far	5
7.2	What I Learned	5
7.3	Risks and Mitigation	5
8	Conclusion	5
A	Additional Details	5

1 Introduction

1.1 Background and Motivation

1.2 Problem Statement

1.3 Aim and Objectives

Aim.

Objectives.

-
-
-
-

2 Related Work

2.1 Model Inversion Attacks

2.2 Split Learning / Edge-Cloud Privacy

2.3 Conditional Entropy Minimization and Surrogates

2.4 Gated Attention Mechanisms

2.5 Slot Attention and Cross-Attention

3 Methodology

3.1 Baseline Overview (CVPR Spotlight)

3.2 Proposed Method 1: Gated-Attention CEM

3.3 Proposed Method 2: Slot + Gated Cross-Attention CEM (Exploratory)

3.4 Evaluation Protocol

4 Implementation

5 Preliminary Results and Analysis

5.1 Main Quantitative Comparison

5.2 Why Method 1 Improves Over the Baseline

5.3 Why Method 2 Underperforms So Far

6 Progress Against Workplan

6.1 Completed Work

-

-
-
-

6.2 Original Plan vs Current Status

6.3 Updated Plan

7 Reflection and Risk Management

7.1 Key Challenges So Far

-
-

7.2 What I Learned

7.3 Risks and Mitigation

- Risk: Mitigation:
- Risk: Mitigation:

8 Conclusion

References

A Additional Details