Qualifying Exam Proposal Outline

Automated Detection and Classification of Cross-Sucking Behavior in Dairy Calves

NSF-Style Structure (June 2026)

PROJECT SUMMARY (1 page)

Overview

- Problem: Cross-sucking affects 60%+ of dairy farms, causes tissue injury
- Gap: No automated detection systems exist
- Solution: Deep learning pipeline for detection and anatomical target classification
- Innovation: Role-conditioned modeling (initiator vs receiver asymmetry)

Intellectual Merit

- Novel RCGAN architecture for asymmetric interaction modeling
- First annotated cross-sucking dataset with role labels
- Addresses extreme class imbalance (87% ear, 12% tail, <1% teat)
- Advances video-based animal behavior understanding

Broader Impacts

- Enables early intervention before tissue damage
- Reduces labor for welfare monitoring
- Transferable to other social behavior detection in livestock
- Open dataset and code for research community

C.1 MOTIVATION AND OBJECTIVES (2-3 pages)

1.1 Background and Problem Statement

- Cross-sucking definition and welfare implications
- Prevalence statistics (cite literature)
- Current monitoring limitations (manual observation impractical)

- Economic impact (teat damage → milk production loss)

1.2 Motivational Scenario

Adapt from Krishnan's "Community Cyber Incident" style

Scenario: A Day on a Modern Dairy Farm

- Describe a realistic scenario where cross-sucking goes undetected
- Multiple pens, hundreds of calves, limited staff
- By the time behavior is noticed, tissue damage has occurred
- Illustrate need for automated, continuous monitoring

1.3 Research Gap Analysis

Current State	Gap	Proposed Solution
Manual observation	Subjective, labor-intensive	Automated detection
Binary detection	No target classification	Ear/tail/teat classification
Treat animals equally	Ignores initiator/receiver roles	Role-conditioned modeling
Balanced datasets	Real data is imbalanced	Novel sampling strategies

1.4 Specific Research Objectives

Objective 1: Develop a validated cross-sucking detection system

- Detect cross-sucking events from overhead video
- Achieve >80% recall on minority classes

Objective 2: Classify anatomical targets (ear/tail/teat)

- Distinguish welfare-critical teat events
- Handle extreme class imbalance

Objective 3: Model initiator-receiver interaction asymmetry

- Design Role-Conditioned Graph Attention Network (RCGAN)
- Leverage finding that initiators control 57% of terminations

Objective 4: Create benchmark dataset and evaluation protocol

- Annotated dataset with bounding boxes and role labels
- Standardized train/val/test splits
- Benchmark for future research

1.5 Hypotheses

H1: Deep learning models can reliably detect cross-sucking events from overhead video with >80% accuracy.

H2: Temporal information (video vs single frame) improves anatomical target classification.

H3: Role-conditioned modeling that treats initiator and receiver separately outperforms symmetric approaches.

H4: Graph-based interaction modeling captures spatial relationships better than independent processing.

1.6 PI Qualifications / Student Background

- Relevant coursework
- Prior research experience
- Preliminary results achieved
- Advisor expertise

C.2 TECHNICAL APPROACH (2-3 pages)

2.1 System Overview

```
┌──────────────────────────────────────────────────────────────┐
│                DETECTION PIPELINE                            │
├──────────────────────────────────────────────────────────────┤
│                                                              │
│ Stage 0: Data Foundation                                     │
│  ├── Raw video → Event extraction                            │
│  ├── Annotation (bboxes, roles, targets)                     │
│  └── Train/Val/Test splits (no leakage)                      │
│                                                              │
│ Stage 1: Calf Detection & Tracking                           │
│  ├── YOLOv8 detector (bootstrapped)                          │
│  └── ByteTrack for temporal consistency                      │
│                                                              │
│ Stage 2: Feature Extraction                                  │
│  ├── Video backbone (R3D-18 / X3D-M / SlowFast)              │
│  └── Per-calf appearance + motion features                   │
│                                                              │
```

```
│  Stage 3: Role-Conditioned Graph Attention (RCGAN)              │
│  ├── Initiator encoder                                          │
│  ├── Receiver encoder                                           │
│  ├── Spatial relationship modeling                              │
│  └── Asymmetric attention mechanism                             │
│                                                                 │
│  Stage 4: Classification                                        │
│  ├── Binary: cross-sucking yes/no                               │
│  └── Target: ear / tail / teat                                  │
│                                                                 │
└─────────────────────────────────────────────────────────────────┘
```

2.2 Role-Conditioned Graph Attention Network (RCGAN)

Key Insight: Initiators control event termination 57% of the time → asymmetric modeling justified

Architecture:

```
Input: Video clip with labeled initiator (I) and receiver (R)

1. Crop tracklets using bounding boxes
   - I_crop = crop(video, initiator_bbox)
   - R_crop = crop(video, receiver_bbox)

2. Extract features with shared backbone
   - f_I = Backbone(I_crop)
   - f_R = Backbone(R_crop)

3. Role-specific encoding
   - z_I = InitiatorEncoder(f_I)  # Focus on head/mouth
   - z_R = ReceiverEncoder(f_R)   # Focus on body region

4. Graph attention
   - Build graph: nodes = {I, R}, edges = spatial relations
   - Apply GAT with role-conditioned attention weights
   - z_interaction = GAT(z_I, z_R, spatial_features)

5. Classification
   - y_target = Classifier(z_interaction)
```

Why Graph Attention?

- Captures spatial relationship (where is mouth relative to body?)
- Handles variable number of calves in frame
- Explicitly models interaction, not just co-occurrence

2.3 Handling Class Imbalance

Strategy	Description	When to Use
Focal Loss	Down-weight easy examples	All training
Class-balanced sampling	Oversample rare classes	Minority <10%
Data augmentation	Synthetic variations	Limited data
Two-stage training	Detect first, then classify	If detection works

2.4 Theoretical Underpinnings

Like Krishnan's use of temporal logic, ground your approach in theory:

- **Information theory:** Mutual information between initiator features and target
- **Graph neural networks:** Message passing for relational reasoning
- **Attention mechanisms:** Interpretable importance weighting

---

C.3 RELATED WORK (2 pages)

3.1 Animal Behavior Recognition

- Livestock monitoring systems
- Cattle behavior detection (cite 0.856 mAP paper)
- Social interaction modeling in animals

3.2 Video Action Recognition

- Two-stream networks
- 3D CNNs (I3D, R3D, SlowFast)
- Video transformers

3.3 Graph Neural Networks for Interaction

- Social interaction recognition
- Group activity recognition
- Human-object interaction

3.4 Class Imbalance in Deep Learning

- Long-tail recognition
- Focal loss, class-balanced loss
- Few-shot and zero-shot learning

3.5 Cross-Sucking Literature (Domain)

- Prevalence and causes
- Welfare implications
- Current intervention strategies

Gap Statement: "No prior work combines (1) automated cross-sucking detection, (2) role-aware asymmetric modeling, (3) graph-based interaction representation, and (4) solutions for extreme class imbalance in this domain."

C.4 PROPOSED RESEARCH (5-6 pages)

Research Task Overview

Task	Description	Timeline
T1	Data Foundation & Annotation	Months 1-3
T2	Baseline Video Classification	Months 2-4
T3	Detection & Tracking Pipeline	Months 3-5
T4	RCGAN Development	Months 4-7
T5	Class Imbalance Strategies	Months 5-8
T6	Evaluation & Ablation	Months 7-9

T1: Data Foundation & Annotation

T1.1: Dataset Curation

- Source: ~1,680 hours of overhead video from 6 calf groups

- Events: ~970 labeled cross-sucking events
- Class distribution: 87% ear, 12% tail, <1% teat

T1.2: Enhanced Annotation (CVAT)

- Bounding boxes for initiator and receiver
- Role labels (initiator/receiver/other)
- Contact point annotation
- Event visibility verification

T1.3: Split Strategy

- Intra-video split for Group 1 (training)
- Group 2 held out for OOD evaluation
- No data leakage verification

Deliverable: Annotated dataset with verified splits

T2: Baseline Video Classification

T2.1: Establish Baselines

- R3D-18 (current)
- X3D-M (efficient)
- SlowFast (dual pathway)
- 2D CNN + LSTM (simpler)

T2.2: Training Protocol

- Focal loss (γ=2)
- Class-balanced sampling
- Strong augmentation (RandAug)
- AMP training for efficiency

T2.3: Evaluation Metrics

- Overall accuracy
- Per-class recall/precision/F1
- Tail-class PR-AUC

- Calibration (ECE)

Deliverable: Baseline performance benchmark

T3: Detection & Tracking Pipeline

T3.1: Calf Detector

- Bootstrap YOLOv8n from CVAT annotations
- Fine-tune on calf detection
- Target: mAP50 > 0.85

T3.2: Tracking

- ByteTrack for multi-object tracking
- Associate detections across frames
- Generate tracklets for each calf

T3.3: Tube Proposals

- Crop initiator/receiver tubes
- Align temporally
- Feed to RCGAN

Deliverable: Detection + tracking pipeline

T4: RCGAN Development

T4.1: Architecture Design

- Role-specific encoders
- Graph construction (spatial, temporal edges)
- Attention mechanism design

T4.2: Training Strategy

- Pre-train encoders on classification
- Fine-tune with graph attention
- Multi-task learning (detection + classification)

T4.3: Interpretability

- Attention visualization
- What does the model focus on?
- Validate with domain knowledge

Deliverable: RCGAN implementation + analysis

T5: Class Imbalance Strategies

T5.1: Loss Function Design

- Focal loss variants
- Class-balanced loss
- Distribution-balanced loss

T5.2: Sampling Strategies

- Oversampling rare classes
- Undersampling majority
- Curriculum learning

T5.3: Data Augmentation

- Temporal augmentation
- Spatial augmentation
- Mixup / CutMix for video

T5.4: Alternative Framing

- Anomaly detection (teat as anomaly)
- Hierarchical classification
- Few-shot learning

Deliverable: Best practices for extreme imbalance

T6: Evaluation & Ablation Studies

T6.1: Main Evaluation

- Compare RCGAN vs baselines
- Statistical significance testing
- Cross-group generalization

T6.2: Ablation Studies

- Role conditioning: symmetric vs asymmetric
- Graph structure: with/without spatial edges
- Backbone: which features matter?

T6.3: Error Analysis

- Confusion patterns
- Failure case analysis
- Calibration assessment

Deliverable: Comprehensive evaluation report

C.5 EVALUATION PLAN (1 page)

5.1 Metrics

Metric	Purpose	Target
Overall Accuracy	General performance	>85%
Tail Recall	Rare class detection	>60%
Teat Recall	Welfare-critical class	>50%
Macro F1	Balanced performance	>0.65
PR-AUC (tail)	Ranking quality	>0.50
ECE	Calibration	<0.10

5.2 Evaluation Protocol

Following EPIC-style reporting:

1. Overall metrics (accuracy, macro F1)

2. Tail-class subset (F1, PR-AUC, recall@precision)
3. Unseen-group generalization (Group 2)

5.3 Statistical Rigor

- Bootstrap confidence intervals (n=1000)
- Multiple random seeds (n=5)
- Significance testing for comparisons

5.4 Baselines for Comparison

- Random baseline
- Majority class baseline
- Frame-level CNN
- Video CNN (R3D, X3D)
- RCGAN (proposed)

C.6 TIMELINE (Gantt Chart)

Task	M1	M2	M3	M4	M5	M6	M7	M8	M9
T1: Data & Annotation	XX	XX	XX						
T2: Baseline Models		XX	XX	XX					
T3: Detection/Tracking			XX	XX	XX				
T4: RCGAN Development				XX	XX	XX	XX		
T5: Class Imbalance					XX	XX	XX	XX	
T6: Evaluation							XX	XX	XX
Writing & Defense								XX	XX

Milestones:

- M3: Annotation complete, baseline established
- M5: Detection pipeline working
- M7: RCGAN implemented
- M9: Evaluation complete, defense ready

C.7 BROADER IMPACTS (0.5 page)

Scientific Impact

- First automated cross-sucking detection system
- Benchmark dataset for livestock behavior research
- Novel approach to asymmetric interaction modeling

Agricultural Impact

- Enables scalable welfare monitoring
- Early intervention prevents tissue damage
- Reduces labor costs for farmers

Methodological Contributions

- Solutions for extreme class imbalance
- Role-conditioned graph attention design
- Transfer potential to other livestock behaviors

Educational Impact

- Open-source code and dataset
- Documentation for practitioners
- Potential for extension to commercial systems

C.8 FIVE CONTRIBUTIONS (for Quals)

1. **Annotated Dataset:** First cross-sucking dataset with bounding boxes, role labels, and verified temporal annotations
2. **Role-Conditioned Architecture:** RCGAN that models initiator-receiver asymmetry for interaction recognition
3. **Extreme Imbalance Solutions:** Evaluated strategies for <1% minority class in video classification
4. **Evaluation Protocol:** EPIC-style reporting with tail-class focus and OOD generalization testing
5. **Practical System:** End-to-end pipeline from raw video to welfare-relevant classifications

---

REFERENCES (40+ papers)

To gather (by category):

Animal Behavior Recognition (10)

- Cattle behavior detection
- Livestock monitoring systems
- Automated welfare assessment

Video Understanding (10)

- 3D CNNs (I3D, R3D, SlowFast, X3D)
- Video transformers
- Temporal modeling

Graph Neural Networks (8)

- GAT, GCN fundamentals
- Interaction recognition
- Social behavior modeling

Class Imbalance (7)

- Focal loss
- Long-tail recognition
- Few-shot learning

Cross-Sucking Domain (5+)

- Prevalence studies
- Welfare implications
- Intervention strategies

APPENDIX: Key Figures to Create

1. **System Architecture Diagram** (like Figure 2 in Krishnan)
2. **RCGAN Architecture Detail**
3. **Dataset Statistics** (class distribution, temporal distribution)

4. **Preliminary Results Table**
5. **Timeline Gantt Chart**
6. **Example Annotations** (bbox + role visualization)