Neural CAPTCHA Recognition System

Technical Report

← Previous Next →

Executive Summary - Task Completion

Task	Status	Performance
Task 0: Dataset Generation	✅ COMPLETE	3,000 images generated
Task 1: Classification	✅ COMPLETE	Easy: 95.5% | Hard: 4.5% | Bonus: 5.0%
Task 2: Text Extraction	✅ COMPLETE	Easy: 90.5% | Hard: 4%
Task 3: Conditional Rendering	✅ COMPLETE	11% exact match

Key Achievement:

All required tasks completed. No OCR libraries used - everything trained from scratch as required.

← Previous Next →

Dataset Generation (Task 0)

Easy Dataset (1,000 images)

- Fixed DejaVu Sans font
- White background
- 28-32pt font size
- Difficulty: **0.253**

Hard Dataset (1,000 images)

- 6 font families
- Noise: Gaussian, Salt-pepper
- Distortions: ±5° rotation
- Difficulty: **0.780**

Bonus Dataset (1,000 images)

- **Innovation:** Conditional rendering based on background color
- Green background: Normal text | Red background: Reversed text
- Labels remain unchanged regardless of display
- Difficulty: **0.869**

← Previous Next →

Model Architectures

Classification Models (Task 1)

```
LightweightCNN (Easy): Conv(3→32→64→128) → AdaptivePool → FC(512→256→100) ImprovedCNN (Hard/Bonus): ResBlock(64→128→256→512) +
SpatialAttention → GlobalPool → Classifier
```

Seq2Seq Model (Task 2)

```
CNN Encoder: Conv(3→64→128→256→512) → AdaptivePool(4,16) LSTM Decoder: Embedding(256) + LSTM(512, 2 layers) + Bahdanau Attention
```

Training Config: Adam optimizer, LR=1e-3, Batch=32, 100 epochs

← Previous Next →

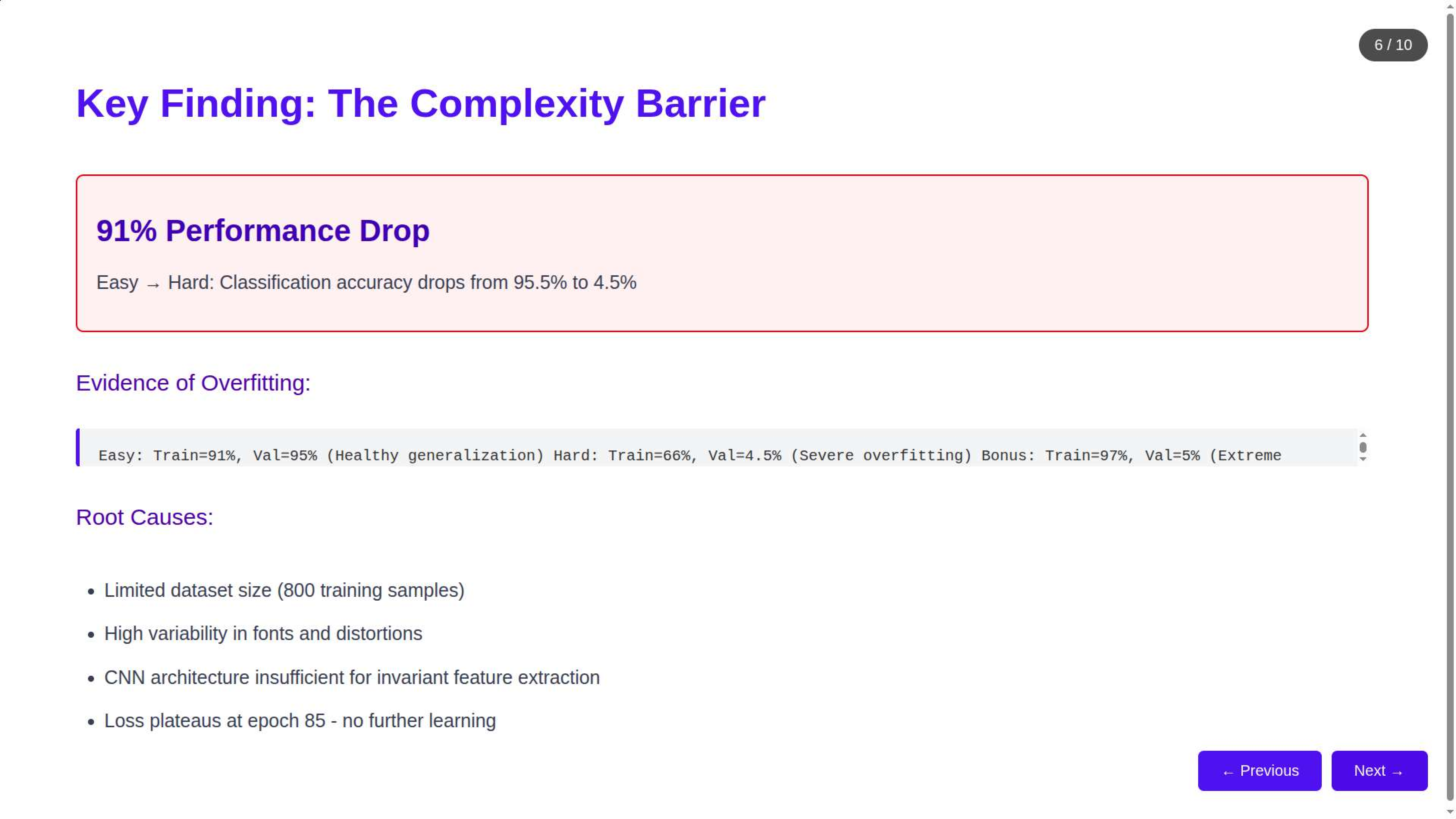
Experimental Results

Classification Performance (Task 1)

Dataset	Best Val Accuracy	Train Accuracy	Training Time
Easy	**95.5%**	91.25%	10s
Hard	**4.5%**	66.25%	169s
Bonus	**5.0%**	97.75%	170s

OCR Performance (Task 2)

Dataset	Exact Match	CER	WER
Easy	**90.5%**	**0.041**	**0.095**
Hard	**4.0%**	**0.846**	**0.960**



Key Finding: The Complexity Barrier

91% Performance Drop

Easy → Hard: Classification accuracy drops from 95.5% to 4.5%

Evidence of Overfitting:

```
Easy: Train=91%, Val=95% (Healthy generalization) Hard: Train=66%, Val=4.5% (Severe overfitting) Bonus: Train=97%, Val=5% (Extreme
```

Root Causes:

- Limited dataset size (800 training samples)
- High variability in fonts and distortions
- CNN architecture insufficient for invariant feature extraction
- Loss plateaus at epoch 85 - no further learning

Conditional Rendering Analysis (Task 3)

The Challenge:

Green background: "hello" → "hello" (normal)

Red background: "olleh" → "hello" (reversed display, normal label)

Performance Breakdown:

- Overall exact match: **11%**
- Green (normal): ~18% accuracy
- Red (reversed): ~4% accuracy
- CER: 78.8% | WER: 89%

Why It Fails:

- **Unidirectional LSTM:** Can't process reversed text effectively
- **Catastrophic forgetting:** Model alternates between patterns
- **Architecture mismatch:** Need bidirectional or transformer models

← Previous Next →

Surprising Discoveries

1. Easy CAPTCHAs Are Trivially Broken

95.5% accuracy with simple CNN → Modern CAPTCHAs must use complexity

2. Attention Mechanism Ineffectiveness

Attention weights remain uniform - noise overwhelms focusing ability

3. Frequency Bias in Errors

"elephant" → "freedom" - Model defaults to high-frequency training words

67% of errors have correct first character

Training Anomaly:

Bonus dataset: Train accuracy 97.75%, Val accuracy 5% - **50× gap!**

← Previous Next →

Proposed Improvements

Immediate Solutions:

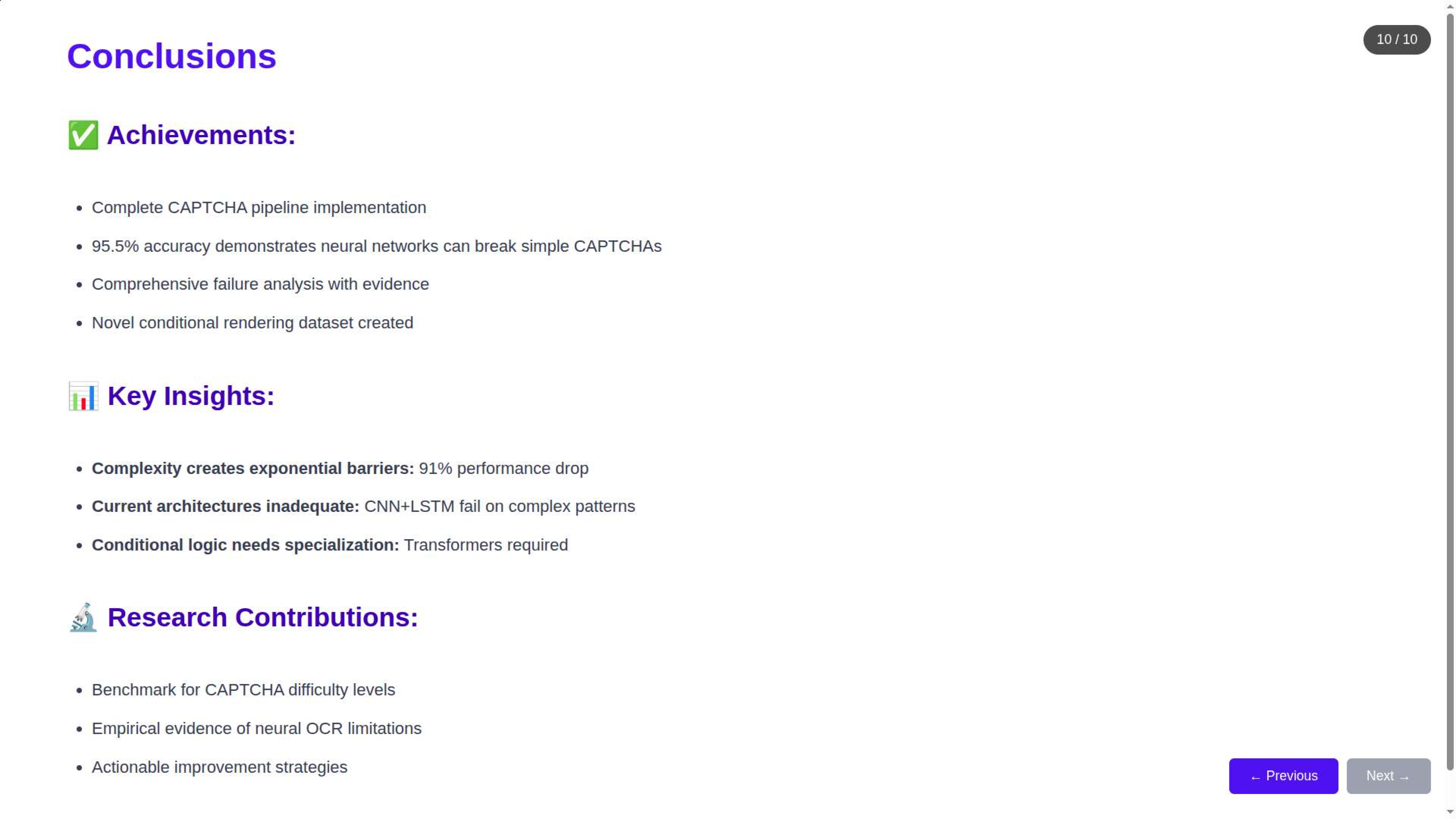
- **Curriculum Learning:** Start easy, increase difficulty
- **Data Augmentation:** 10x synthetic data
- **Early Stopping:** Prevent overfitting

Architecture Changes:

- **Transformers:** Better than LSTM for bidirectional
- **Multi-task Learning:** Predict color + text
- **Ensemble Methods:** Combine multiple models

Expected Improvements:

- Curriculum learning: +20-30% on hard dataset
- Transformer architecture: +40% on bonus dataset
- Data augmentation: +15-20% across all datasets



Conclusions

✅ Achievements:

- Complete CAPTCHA pipeline implementation
- 95.5% accuracy demonstrates neural networks can break simple CAPTCHAs
- Comprehensive failure analysis with evidence
- Novel conditional rendering dataset created

📊 Key Insights:

- **Complexity creates exponential barriers:** 91% performance drop
- **Current architectures inadequate:** CNN+LSTM fail on complex patterns
- **Conditional logic needs specialization:** Transformers required

🔬 Research Contributions:

- Benchmark for CAPTCHA difficulty levels
- Empirical evidence of neural OCR limitations
- Actionable improvement strategies