# Breaking the Parameter Scaling Barrier in Deep RL: The Power of Soft Mixtures of Experts and Tokenization

Based on Ceron et al., 2024[1] and Sokar, Obando-Ceron, et al., 2024[2]

Beijing Institute for General Artificial Intelligence (BIGAI)

May 23, 2025

[1] Johan S Obando Ceron et al. (2024). "Mixtures of Experts Unlock Parameter Scaling for Deep RL". In: *International Conference on Machine Learning*

[2] Ghada Sokar, Johan S Obando-Ceron, et al. (2024). "Don't flatten, tokenize! Unlocking the key to SoftMoE's efficacy in deep RL". In: *arXiv preprint arXiv:2408.13637*

# Two Groundbreaking Studies, One Revolutionary Insight

## Paper 1: Ceron et al. (2024)

**Finding:** SoftMoE enables parameter scaling
**Question:** Why does it work?

- ✓ Proved scaling is possible
- ✓ Showed consistent improvements
- ? Mechanism unclear

## Paper 2: Sokar et al. (2024)

**Finding:** Tokenization is the key
**Answer:** It's not about experts!

- ✓ Isolated the mechanism
- ✓ Single expert works
- ✓ Tokenization critical

**Combined Insight:** The secret is not multiple experts, but how we structure the data flow through tokenization

# The Parameter Scaling Problem in Deep RL

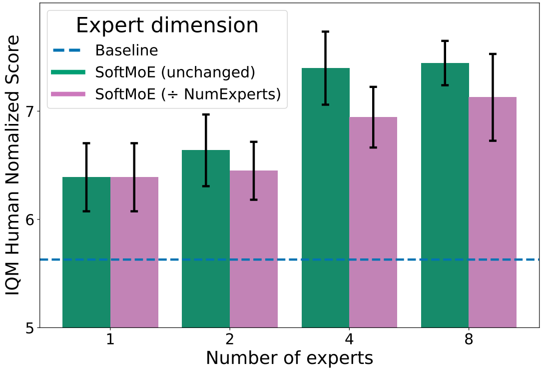


Figure: Traditional approach hurts performance

## Key Challenges:

- Adding parameters **hurts** RL performance Ceron et al., 2024[a]
- Difficult to develop scaling laws Kaplan et al., 2020[b]
- Need new architectural approaches

[a] Johan S Obando Ceron et al. (2024). "Mixtures of Experts Unlock Parameter Scaling for Deep RL". In: *International Conference on Machine Learning*

[b] Jared Kaplan et al. (2020). "Scaling laws for neural language models". In: *arXiv preprint arXiv:2001.08361*

# Why Traditional Scaling Methods Fail


Figure: Unique phenomena in deep RL

## Key Challenges:

- Dormant neurons Sokar, Agarwal, et al., 2023[a]
- Passive learning difficulties
- Capacity loss issues
- Regularization needs Kumar et al., 2021[b]

[a] Ghada Sokar, Rishabh Agarwal, et al. (2023). "The dormant neuron phenomenon in deep reinforcement learning". In: *International Conference on Machine Learning*, pp. 32145–32168

[b] Aviral Kumar et al. (2021). "Dr3: Value-based deep reinforcement learning requires explicit regularization". In: *International Conference on Learning Representations*

# Soft Mixture of Experts (SoftMoE) Architecture

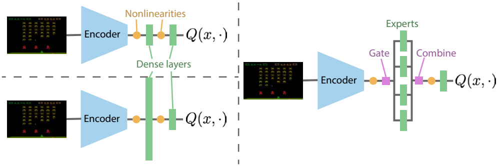


Figure: SoftMoE integration in deep RL

**Key Features:**

- **Soft Assignment**: Differentiable routing
- **Strategic Placement**: Replaces dense layer
- **Spatial Preservation**: Maintains structure
- No discrete routing decisions
- Fully end-to-end trainable
- Computationally efficient

# SoftMoE's Breakthrough Results

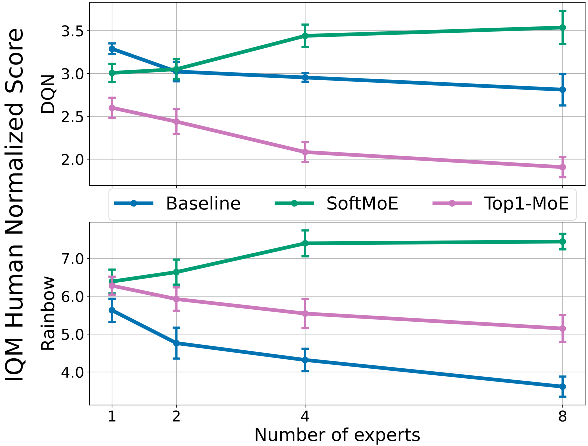


Figure: SoftMoE achieves parameter scalability

**Key Results:**

- **20%** improvement with 8 experts
- **40%** degradation with traditional scaling
- Robust across different settings Ceron et al., 2024[a]

[a] Johan S Obando Ceron et al. (2024). "Mixtures of Experts Unlock Parameter Scaling for Deep RL". In: *International Conference on Machine Learning*

# Tokenization Scheme Comparison

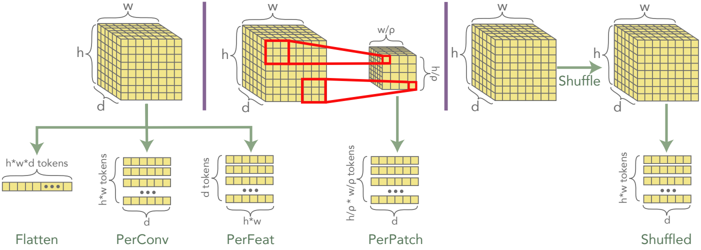


Figure: Architectural comparison of different tokenization strategies

- **PerConv**: $(h, w, d) \rightarrow h \times w$ tokens of dimension $d$
- **PerFeat**: $(h, w, d) \rightarrow d$ tokens of dimension $h \times w$
- **PerSamp**: Entire output as single token

# Surprising Discovery: Single Expert Can Succeed

**SHOCKING REVELATION**

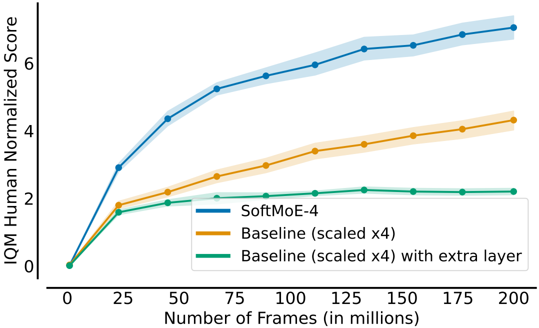


Figure: Single expert vs multi-expert comparison

**Core Insight**

**Tokenization**, not multiple experts, drives SoftMoE's success!

This challenges the assumption that expert diversity drives performance.

# Tokenization Effectiveness Validation

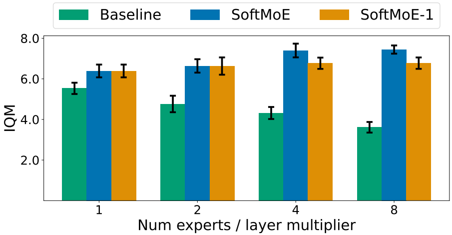


Figure: Tokenized baseline vs traditional baseline

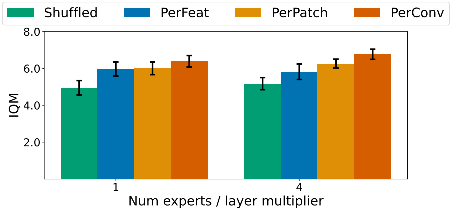


Figure: Effects of different tokenization schemes

- Simple tokenized baseline significantly improves performance
- PerConv tokenization works best
- Preserving spatial structure is crucial

# Architecture Comparison: Traditional vs Tokenized

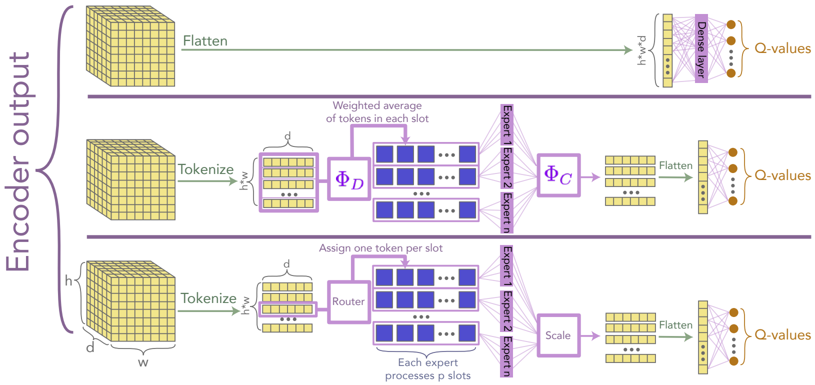


Figure: Baseline, SoftMoE, and Top-k MoE architecture comparison

- **Traditional approach**: Flattens encoder output
- **MoE approach**: Tokenizes then processes by experts
- **Key difference**: Preserving vs losing spatial information

# Expert Utilization Analysis: The Redundancy Problem

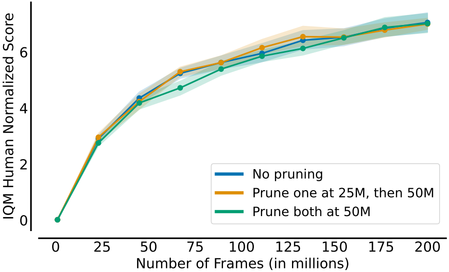


Figure: Expert pruning experiment

## Pruning Test Results

Remove 50% of experts:

- Performance drop: $< 2\%$
- Training: **Stable**
- Convergence: **Unchanged**

## Implications

- Experts learn similar representations
- Scaling experts $\neq$ Scaling performance
- Tokenization does the heavy lifting

**Paradigm Shift:** Focus on tokenization design, not expert

# Cross-Algorithm Effectiveness

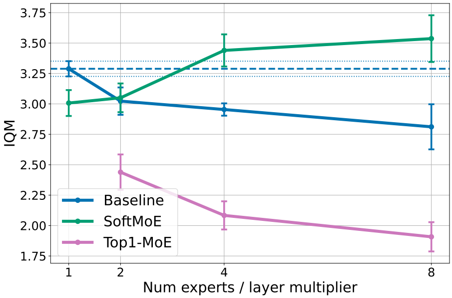


Figure: DQN algorithm results

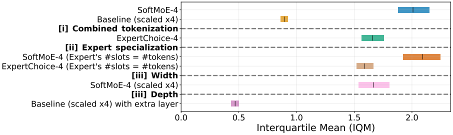


Figure: DER algorithm results

- **DQN**: Basic value function learning
- **Rainbow**: Combination of multiple improvements
- **DER**: Data efficiency regularization method
- Tokenization shows improvements across different algorithms

# Network Architecture Generality

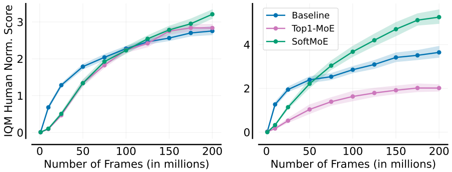


Figure: CNN encoder results

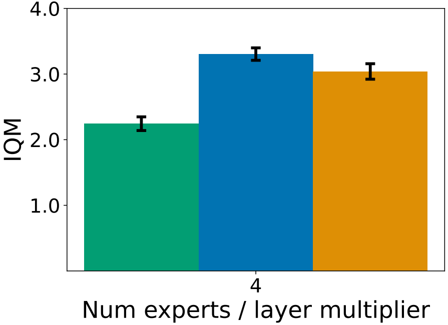


Figure: Full 60-game suite results

- Works not only with Impala but also standard CNN encoders
- Improvements across all 60 Atari games
- Method has broad applicability

# Rethinking Deep RL Architecture Design

## Core Insights

- **Flattening is harmful**: Traditional flattening may lose critical spatial information
- **Tokenization is key**: Spatial structure-preserving tokenization drives performance gains
- **Expert redundancy**: Current MoE setups have low expert utilization
- **Parameter scaling**: Correct architecture design enables effective parameter scaling

## Design Principles

- Prioritize spatial information preservation
- Explore more effective expert allocation strategies
- Reevaluate traditional architectural assumptions

# Future Research Directions

**Architecture Optimization**

- More efficient tokenization schemes
- Expert specialization mechanisms
- Adaptive routing strategies
- Computational efficiency optimization

**Application Extension**

- Continuous control tasks
- Multi-task reinforcement learning
- Offline reinforcement learning
- Large-scale environments

**Theoretical Understanding**

- Theoretical foundations of tokenization
- Expert dynamics analysis
- Scaling law research
- Generalization capability analysis

**Practical Deployment**

- Distributed training optimization
- Inference efficiency improvement
- Hardware adaptation
- Industrial applications

## Research Summary

# Complete Path from Problem to Solution

1. **Problem Identification**: Parameter scaling difficulties in deep RL Ceron et al., 2024[3]
2. **Solution**: SoftMoE achieves parameter scalability
3. **Mechanism Discovery**: Tokenization is the key success factor Sokar, Obando-Ceron, et al., 2024[4]
4. **Deep Insights**: Rethinking architectural design paradigms

### Main Contributions

- First effective parameter scaling in deep RL
- Revealed tokenization's key role in MoE success
- Provided new guidance principles for future architecture design

[3] Johan S Obando Ceron et al. (2024). "Mixtures of Experts Unlock Parameter Scaling for Deep RL". In: *International Conference on Machine Learning*

[4] Ghada Sokar, Johan S Obando-Ceron, et al. (2024). "Don't flatten, tokenize! Unlocking the key to SoftMoE's efficacy in deep RL". In: *arXiv preprint arXiv:2408.13637*

## Key Takeaways

**What We Learned:**

- Traditional flattening loses spatial information
- Tokenization preserves crucial structure
- Single expert + tokenization $\approx$ Multi-expert performance
- Expert redundancy is a real issue

**Design Guidelines:**

- Preserve spatial structure in conv outputs
- Focus on tokenization over expert count
- Consider computational efficiency
- Evaluate expert utilization

**The paradigm shift:** From "How many experts?" to "How to tokenize?"

# Impact & Broader Implications

## Immediate Impact

- Enables parameter scaling in deep RL for the first time
- Works across multiple RL algorithms (DQN, Rainbow, DER)
- Applicable to different network architectures (CNN, Impala)
- Validated on large-scale benchmarks (60 Atari games)

## Broader Implications

- Challenges conventional architectural wisdom in RL
- Opens new research directions in tokenization schemes
- Potential applications beyond visual RL tasks
- Foundation for developing RL scaling laws

# Thank You!

Questions & Discussion

*Rethinking Deep RL Architecture Design:*
*From Flattening to Tokenization*