Transferability of Sparse Subnetworks in Deep Reinforcement Learning

Sebastian Pusch[1] (s5488079), Mihai Damian[1] (s5508983), Yitong Lin Yang[1] (s5460611), Riandika Marshequila Dinu[1] (s5603838)
{s.pusch,m.damian,y.lin.39,riandika.marshequila.dinu}@student.rug.nl

[1]**University of Groningen - Reinforcement Learning Practical - Group 9**

1 Motivations

Modern neural networks in supervised learning are typically heavily overparameterized, a phenomenon already noted in classic work such as Optimal Brain Damage (LeCun et al., 1990), which showed that many parameters can be removed with little loss in performance. The Lottery Ticket Hypothesis (LTH) (Frankle & Carbin, 2019) uses iterative magnitude pruning to identify sparse subnetworks—“winning tickets”—inside dense, randomly initialized networks, then resets their surviving weights to their original values and retrains them in isolation. These subnetworks can match or exceed the accuracy of the full model, often training faster and more reliably, and follow-up work shows that they can sometimes transfer across related datasets (Ganesh et al., 2023), suggesting that they encode meaningful inductive biases rather than merely overfitting to a single task.

In deep reinforcement learning, transfer has been far less successful. Sabatelli and Geurts (2021) examine transfer learning in Atari by pre-training a dense Deep Q-Network (DQN) on a source game and transferring the convolutional encoder to a target game while reinitializing the final value layer. They report limited or negative transfer and conjecture that the tight co-evolution of the feature extractor and value function approximator during training causes negative transfer: reinitializing the head breaks this coupling, leaving the pretrained features detached from the new value function. This highlights the importance of preserving representational alignment during transfer.

At the same time, recent work shows that sparsity itself can be effective in reinforcement learning (RL) when handled appropriately. Obando-Ceron et al. (2023) demonstrate that pruned value-based agents can match or surpass dense performance even at high sparsity levels, indicating that significant parameter reductions do not necessarily degrade policy quality. Todorov et al. (2025) further find that dynamically sparse methods such as Sparse Evolutionary Training (SET) and Gradual Magnitude Pruning (GMP) are more adaptive and robust in multi-task RL than both dense agents and static sparse models. Together, these results suggest that sparse RL networks can perform well when sparsity is applied in a manner that preserves learning dynamics.

Despite these encouraging results, it remains unclear how sparsity interacts with transfer. Prior work has either examined transfer in dense agents (Sabatelli & Geurts, 2021) or studied sparsity in isolation (Obando-Ceron et al., 2023; Todorov et al., 2025), without considering whether sparse subnetworks identified on one RL task—such as winning tickets—can be effectively reused on another. In particular, the role of dynamic sparsification during adaptation to a new task has not been explored. Understanding this interaction is key to determining whether transferable sparse representations can be constructed in RL.

2 Research Question

To what extent do sparse subnetworks ("winning tickets") discovered on a source reinforcement learning task transfer to related target tasks, and how does introducing dynamic sparsification during target-task training (e.g., SET or GMP) affect transfer performance and network plasticity when compared against dense baselines and static sparse tickets?

3 Methodology

We investigate our research question using Double Deep Q-Networks (DDQN) and Deep Quality-Value Networks (DQV), two standard value-based deep RL algorithms commonly used in transfer-learning studies, on a subset of the Atari benchmark. Three environments are selected from the Atari benchmark (Pong, Breakout, and Space Invaders), representing related visual control tasks with differing transition dynamics. Each environment serves once as a source and once as a target, yielding nine source–target combinations including self-transfer. All agents share the same architecture, which consists of a standard convolutional encoder and a value head. Sparsification during training is performed using GMP and SET. All models, pruning procedures, and training loops are implemented in PyTorch, using standard experience replay and target-network updates.

3.1 Phase 1: Finding Sparse Subnetworks on the Source Task

To establish reliable reference performance, we first tune dense DDQN and DQV agents on each environment. We then identify sparse subnetworks using two magnitude-based mask-finding procedures:

Iterative Magnitude Pruning (IMP). Following the Lottery Ticket Hypothesis framework, the dense network is:

1. trained to convergence,
2. pruned by removing a fraction of the lowest-magnitude weights,
3. rewound to its original initialization.

Repeating this process produces a sparse mask M_{IMP} at a target sparsity.

Gradual Magnitude Pruning (GMP). We also apply the gradual schedule of Zhu & Gupta (2017), which increases sparsity from s_0 to s_T between pruning steps t_0 and t_1:

$$s(t) = s_T + (s_0 - s_T)\left(1 - \frac{t - t_0}{t_1 - t_0}\right)^3 .$$

Here, $s(t)$ is the sparsity at pruning step t, s_0 is the initial sparsity, and s_T is the final target sparsity. The parameters t_0 and t_1 mark the start and end of the pruning schedule, and the cubic term controls the rate at which sparsity increases. At each pruning event, the lowest-magnitude weights are removed, yielding a final sparse mask M_{GMP}.

Moderate sparsity levels (60–80%) are used to retain capacity for re-alignment with the newly initialized value head, consistent with the conjecture of Sabatelli & Geurts (2021) that highly specialized feature extractors hinder transfer.

Baselines, tuning, and ablations. All experiments use PyTorch and standard Atari training protocols (Mnih et al., 2015). Dense-from-scratch DDQN/DQV agents act as baselines. In Phase 1 we tune learning rate, target-update frequency, and pruning parameters, and perform ablations over pruning schedules and sparsity levels to ensure stable mask discovery.

We perform controlled ablation by varying one component at a time while keeping all other elements fixed:

(a) Mask source (IMP vs. GMP): We fix all optimization settings, architectures, and sparsification hyperparameters, and vary only the mask-finding procedure used on the source task. Specifically, we compare subnetworks obtained via IMP against those obtained via GMP. This ablation isolates whether transfer behaviour is determined primarily by the type of sparse mask discovered during source-task training.

(b) Sparsity level (60%, 70%, 80%): We fix the pruning schedule, learning rate, and target-network update frequency, and vary only the final sparsity level of the subnetworks. By evaluating masks at 60%, 70%, and 80% sparsity, this ablation tests how sensitive transfer performance and plasticity metrics are to the amount of parameter removal, independent of mask source or adaptation strategy.

We base our tuning choices on prior work showing that sparse deep RL agents are highly sensitive to optimization settings. Learning rate and target-network update frequency are the two hyperparameters with the strongest effect on stability in pruned value-based agents, mainly because sparsity increases gradient variance (Obando-Ceron et al., 2023). Dynamic sparsification makes this sensitivity even stronger, so pruning-schedule parameters also become important for preserving plasticity (Todorov et al., 2025). For this reason, we limit tuning to these components and follow established recommendations for sparse DQN-style systems (Obando-Ceron et al., 2023; Todorov et al., 2025), ensuring that mask quality is not affected by unstable optimization.

3.2 Phase 2: Transfer and Adaptation on the Target Task

After transferring the IMP- or GMP-based subnetworks to the target task, we vary how sparsity is handled during target-task learning. We compare subnetworks obtained via IMP and GMP on the source task, and we consider three adaptation regimes during target-task training: (i) no further pruning (static sparse ticket), (ii) GMP (dynamic magnitude-based pruning), and (iii) SET (dynamic rewiring at fixed sparsity). This yields six core configurations:

$$\text{IMP-ticket} \rightarrow \{\text{none, GMP, SET}\}, \qquad \text{GMP-ticket} \rightarrow \{\text{none, GMP, SET}\}.$$

We additionally include dense-from-scratch and sparse-only-on-target baselines to separate the effects of transfer from the effects of sparsity. Finally, we perform ablations over dynamic sparsity regimes (none, GMP, SET) and mask sources (IMP vs. GMP) to isolate the contribution of each component to transfer behaviour.

3.3 Metrics

We evaluate performance using the Interquartile Mean (IQM) of episodic return (Agarwal et al., 2021), reported with 95% bootstrapped confidence intervals over ten seeds. To quantify transfer benefit, we compute the *Area Ratio(r)*, defined as

$$r = \frac{\text{Area}_{\text{transfer}} - \text{Area}_{\text{scratch}}}{\text{Area}_{\text{scratch}}},$$

which measures the overall improvement or degradation relative to training from scratch.

3.4 Hypothesis

If our hypothesis is correct, subnetworks identified on the source task will exhibit positive transfer once dynamic sparsity is reintroduced during target-task training. We expect SET-based adaptation to yield the strongest transfer benefits, with both IMP-ticket $\rightarrow$ SET and GMP-ticket $\rightarrow$ SET outperforming static sparse tickets and GMP-only adaptation. Among all transfer configurations, we anticipate that IMP-ticket $\rightarrow$ SET will produce the largest improvement relative to IMP-ticket $\rightarrow$ none. Finally, although dense-from-scratch and sparse-only-on-target agents do not involve transfer, we expect the best-performing transfer configurations to achieve competitive or superior *absolute* performance, as measured by interquartile mean (IQM), compared to these baselines.

References

Rishabh Agarwal, Max Schwarzer, Pablo Samuel Castro, Aaron C. Courville, and Marc G. Bellemare. Deep reinforcement learning at the edge of the statistical precipice. In *Advances in Neural Information Processing Systems*, volume 34, pp. 29304–29320, 2021.

Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. In *International Conference on Learning Representations*, 2019.

P. Ganesh, G. Tirumala, A. Singh, et al. One ticket to win them all: Generalizing lottery ticket initializations across datasets and optimizers. *Advances in Neural Information Processing Systems*, 2023.

Yann LeCun, John Denker, and Sara Solla. Optimal brain damage. *Advances in Neural Information Processing Systems*, 2, 1990.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei Rusu, Joel Veness, Marc G. Bellemere, Alex Graves, Martin Riedmiller, Andreas Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, feb 2015. ISSN 0028-0836. URL http://dx.doi.org/10.1038/nature14236.

Julio Obando-Ceron, Pablo Samuel Castro, and Marc G. Bellemare. Pruning deep reinforcement learning networks without performance loss. In *International Conference on Learning Representations*, 2023.

Matthia Sabatelli and Pierre Geurts. On the transferability of winning tickets in reinforcement learning. *arXiv preprint arXiv:2110.02639*, 2021.

Aleksandar Todorov, Juan Cardenas-Cartagena, Rafael F. Cunha, Marco Zullich, and Matthia Sabatelli. Sparsity-driven plasticity in multi-task reinforcement learning. *Transactions on Machine Learning Research*, 2025.

Michael Zhu and Suyog Gupta. To prune, or not to prune: Exploring the efficacy of pruning for model compression. In *International Conference on Learning Representations (Workshop Track)*, 2017.