Higher Resolution, Better Generalization: Unlocking Visual Scaling in Deep Reinforcement Learning

Anonymous authors
Paper under double-blind review

Abstract

Pixel-based deep reinforcement learning agents are typically trained on heavily downsampled visual observations, a convention inherited from early benchmarks rather than grounded in principled design. In this work, we show that observation resolution is a critical yet overlooked variable for policy learning: higher-resolution inputs can substantially improve both performance and generalization, provided the network architecture can process them effectively. We find that the widely used Impala encoder, which flattens spatial features into a vector, suffers from quadratic parameter growth as resolution increases and fails to leverage the additional visual detail. Replacing this operation with global average pooling, as in the Impoola architecture, decouples parameter count from resolution and yields consistent improvements across resolutions and network widths—at their respective best conditions, visual scaling unlocks a 28 % performance gain for Impoola over Impala. These gains are strongest in environments that require precise perception of small or distant objects, and gradient saliency analysis confirms that the underlying mechanism is a more spatially localized visual attention of the policy at higher resolutions. Our results challenge the prevailing practice of aggressive input downsampling and position resolution-independent architectures as a simple, effective path toward scalable visual deep RL.

1 Introduction

Benchmarks in deep learning go beyond quantifying progress; they actively steer the direction of research itself (Dehghani et al., 2021; Raji et al., 2021; Koch et al., 2021). The structure of benchmarks, e.g., input dimensionality, output format, and data distribution, implicitly determines which architectures are viable and which design choices receive sustained attention. When benchmarks evolve, the methods built around them evolve in response. For instance, the transition in computer vision (CV) from low-resolution datasets like CIFAR-10 (Krizhevsky et al., 2009) to higher-resolution benchmarks like ImageNet (Russakovsky et al., 2015; Deng et al., 2009), was not merely a scaling exercise; it motivated deeper networks (Krizhevsky et al., 2012; Simonyan & Zisserman, 2014; Szegedy et al., 2015), hierarchical feature extractors (Lin et al., 2017), and principled strategies for allocating capacity across layers (Tan & Le, 2019; 2021). As such, resolution is not seen as a superficial preprocessing detail but a first-order constraint that shapes architectural development.

Recognizing this, the CV community has designed architectures that treat resolution as a tunable parameter rather than a fixed constraint. By utilizing global average pooling (GAP) (Lin et al., 2014), models like ResNet (He et al., 2016) and ConvNeXt (Liu et al., 2022) decouple their parameter count from input dimensionality. This flexibility allows models to be pre-trained at standardized scales, e.g., $(224, 224)$, and seamlessly adapted to high-fidelity downstream tasks, e.g., from satellite imagery for land cover classification (Penatti et al., 2015; Helber et al., 2019) to medical diagnostics including chest radiography and dermatology (Rajpurkar et al., 2017; Esteva et al., 2017), without altering the underlying backbone. Research into the train-test resolution discrepancy, such as FixRes (Touvron et al., 2019), further shows that vision models can be fine-tuned across resolutions to improve performance. In CV, resolution has become an adjustable variable while it remains a rigid, largely unexamined constraint in reinforcement learning (RL).

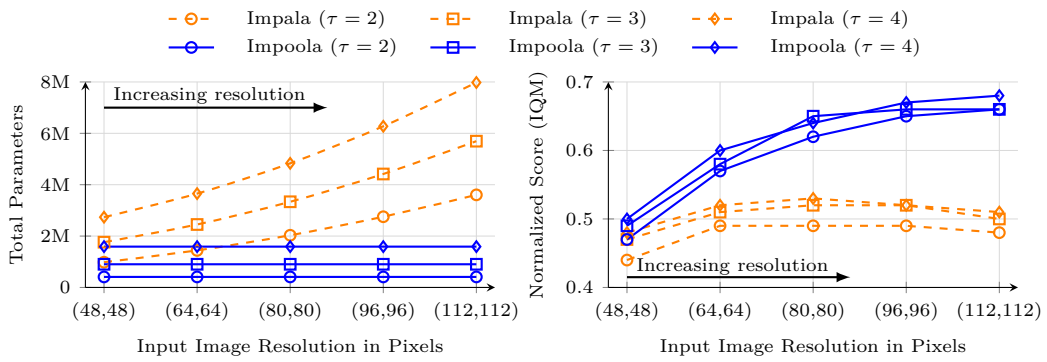


Figure 1: The impact of scaling the input image resolution on the network's total parameter count (**left**) and the performance as normalized score (**right**). We compare the common image encoders Impala (Espeholt et al., 2018) and Impoola (Trumpp et al., 2025) across resolutions from $(48, 48)$ to $(112, 112)$ pixels. Different network widths are shown, i.e., the number of filters per Conv2d layer is scaled by τ. The counts include the parameters for separate actor and critic heads. It can be seen that the parameter count in Impala increases with resolution due to its Flatten layer, whereas Impoola remains constant due to its GAP layer. Given this, we find that Impoola demonstrates superior scaling behavior to Impala, whereas visual scaling unlocks a 28 % performance gain for Impoola over Impala at their respective best conditions. These results highlight our proposition that performance gains can be achieved by scaling input resolutions in deep RL.

This rigidity is consequential because visual observations in deep RL are not merely inputs to be classified, but states upon which sequential decisions are conditioned (Mnih et al., 2013; Levine et al., 2016). Therefore, resolution directly affects state aliasing (Whitehead & Ballard, 1991), partial observability (Hausknecht & Stone, 2015), and the precision required for long-horizon control. When agents are trained and evaluated solely on low-resolution abstractions, they risk developing policies that exploit coarse-grained artifacts rather than learning robust spatial features; a form of shortcut learning (Geirhos et al., 2020; Song et al., 2020; Zhang et al., 2018) that undermines transfer to the high-fidelity inputs encountered in real-world deployment (Dulac-Arnold et al., 2019; Ibarz et al., 2021). Yet widely used deep RL benchmarks, including Atari (Bellemare et al., 2013) and Procgen (Cobbe et al., 2020), rely almost exclusively on low-resolution observations. Historically, this was a pragmatic necessity: the computational cost of simulator throughput and the sample-inefficiency of RL algorithms made high-resolution training prohibitively expensive. Environments are routinely downsampled to coarse grids, e.g., $(84, 84)$ for Atari and $(64, 64)$ for Procgen, discarding spatial information and fixing the perceptual scale at which agents operate (Machado et al., 2018).

Because the field has been anchored to low-resolution inputs, standard architectures have been designed around this assumption. The widely used Impala encoder (Espeholt et al., 2018) relies on a flattening operation to transition from convolutional features to the policy head. When resolution increases, this operation causes the first fully connected layer to grow quadratically with spatial input size, concentrating the vast majority of parameters in a single layer; see Figure 1 (left). The resulting severe capacity allocation asymmetry hinders the model from effectively leveraging higher-dimensional inputs. Thus, the low-resolution paradigm has shaped not only evaluation practices but the architectural foundations of deep RL itself.

As computational resources and simulator efficiency improve, there is an opportunity to revisit these constraints—yet no systematic study has examined how observation resolution interacts with architecture and learning dynamics in deep RL. Among existing benchmarks, Procgen is particularly well-suited for this study: it is open-source, procedurally generated, computationally lightweight, and offers 16 environments with diverse perceptual demands and built-in easy/hard difficulty modes.

Building this study on a configurable-resolution extension of Procgen, our main contributions are as follows:

- We show in a systematic study that observation resolution is a first-order variable for policy learning. Across 16 Procgen environments, five resolutions, three network widths, and multiple training regimes, higher-resolution inputs yield substantial improvement in aggregate performance and gen-

eralization without any algorithmic changes, but only when paired with a resolution-independent architecture as shown in Figure 1 (right). The standard Impala encoder fails to leverage additional visual detail, presumably because its Flatten layer exhibits quadratic parameter growth.

- Our results for the hard training regime show that when sufficient performance headroom remains, higher resolution reliably improves both architectures, including standard Impala. Environment-level analysis reveals that the largest gains consistently occur in environments that require precise perception of small or distant entities, where standard resolutions create a perceptual bottleneck.
- We provide mechanistic evidence for these gains through gradient saliency and dormant neuron analysis. Higher resolution enables more spatially focused policy attention on task-critical entities, while the flatten-based architecture disperses its gradient signal and accumulates inactive neurons as resolution grows.
- Lastly, we introduce Procgen-HD, an open-source extension of the Procgen benchmark that supports arbitrary rendering resolutions while preserving identical game logic, level generation, and reward structure. Procgen-HD enables controlled experiments that isolate the effect of visual fidelity on policy learning, and we release it to facilitate future research on resolution scaling in deep RL.

Our findings challenge the prevailing practice of aggressive input downsampling and suggest that resolution-independent architectures offer a simple, underexplored path to stronger visual deep RL. Subsequent studies may further explore resolution as a tunable variable rather than a fixed constraint, as we show that significant performance and generalization gains are also available without algorithmic innovation.

2 Related Work

Benchmarks and Preprocessing Conventions in Visual Deep RL. The standard (84,84)-pixel grayscale image that dominates visual RL emerged from pragmatic constraints rather than principled design. Mnih et al. (2013; 2015) introduced this preprocessing for DQN, explicitly noting that the cropping was required because their GPU convolution implementation expected square inputs. The original Atari frames of (210,60) with 7-bit colors were downsampled and cropped to (84,84) grayscale, with 4 frames stacked to handle partial observability (e.g direction of ball in the Pong.) and action repeat of 4 to reduce computational overhead. These choices, born from 2013-era hardware limitations, became unexamined standards that persist across modern benchmarks.

Machado et al. (2018) revisited the Arcade Learning Environment (Bellemare et al., 2013) and codified best practices, introducing sticky actions to prevent determinism exploitation. However, this influential work inherited rather than questioned the resolution choice, focusing on stochasticity, episode termination, and evaluation protocols while accepting (84,84) as given. Similarly, Braylan et al. (2015) demonstrated that frame skip significantly affects performance with optimal values varying by game, yet no analogous study examined resolution as a variable.

Subsequent benchmarks inherited these conventions with minor modifications. Procgen (Cobbe et al., 2020) uses (64,64) RGB observations without frame stacking, relying on procedural generation to prevent memorization, but does not justify the resolution choice or provide built-in support for varying it. Similarly, pixel-based DMControl (Tunyasuvunakool et al., 2020) experiments, as standardized by algorithm papers like SAC+AE (Yarats et al., 2021) and DrQ (Yarats et al., 2020), settled on (84,84) RGB with 3-frame stacking, again inheriting from Atari conventions rather than empirical validation. The pattern is consistent: benchmark designers select convenient resolutions that then harden into convention, with reproducibility pressure discouraging deviation.

Scaling in Vision-Based Deep RL. Compute scaling through distributed training increased throughput without revisiting visual preprocessing. IMPALA (Espeholt et al., 2018) introduced actor-learner architectures while establishing the Impala model as the de facto encoder. SEED RL (Espeholt et al., 2020) and R2D2 (Kapturowski et al., 2019) further improved throughput and sample efficiency. All these systems kept resolution fixed at (64,64), focusing on sample throughput rather than input fidelity.

However, architectural scaling in RL differs fundamentally from supervised learning. Several pathologies explain this: dormant neurons accumulate due to target non-stationarity (Sokar et al., 2023), primacy bias causes overfitting to early experiences (Nikishin et al., 2022), and plasticity loss reduces the ability to fit new targets (Lyle et al., 2022; Abbas et al., 2023). Kumar et al. (2021) identified implicit under-parameterization, where value networks experience rank collapse despite nominal capacity.

Recent work has identified strategies that address these pathologies. Ceron et al. (2024) demonstrated that mixture-of-experts unlock parameter scaling by reducing dormant neurons, with Willi et al. (2024) extending these findings to multi-task settings. Network sparsity offers another path: Graesser et al. (2022) showed 90% sparsity matches or exceeds dense baselines, while Sokar et al. (2022) found sparse networks learn faster by avoiding memorization of early samples.

Most relevant to resolution scaling, concurrent work has identified the flatten operation as a critical bottleneck. Sokar et al. (2025) showed that flattening creates a high-dimensional bottleneck that impedes scaling, whereas tokenization preserves spatial structure. Trumpp et al. (2025) introduced the Impoola encoder, replacing flattening with global average pooling (GAP), further reiterated by the parallel work of Sokar & Castro (2025) on GAP, and Kooi et al. (2025) proposed Hadamax encoding with similar benefits. GAP provides natural resolution independence by reducing the spatial dimensions to a single value per channel, enabling networks to handle varying resolutions without parameter explosion. To our knowledge, no prior work has systematically studied the effect of *observation resolution* on learning dynamics in deep RL.

3 Experimental Setup

The goal of this work is to analyse the relationship between performance and increased input resolutions as model capacity increases. We describe our benchmark in the following and provide details on the network architecture and our training settings.

3.1 Benchmark:

This work's analysis is based on the Procgen Benchmark (Cobbe et al., 2020), a suite of 16 procedurally generated environments that provides a more robust evaluation of agent capabilities compared to the fixed environments of the Atari suite. While Atari games rely on a static set of textures and deterministic transitions, Procgen environments feature high-variance procedural generation, ensuring that agents encounter unique visual combinations of layouts and assets in every episode. This structural and visual diversity provides a clearer signal for evaluating the impact of input resolution, as the agent must resolve fine-grained details in constantly changing environments rather than relying on fixed pixel patterns.

To systematically investigate the effects of scaling, we use our custom modification of the benchmark, *Procgen-HD*, which supports configurable rendering resolutions. Formally, let the observation $x \in \mathbb{R}^{R \times R \times 3}$ denote a tensor of an RGB image with (R, R) pixels, i.e., the resolution is $\mathrm{d}(x) = (R, R)$. We evaluate performance across the discrete set of resolutions $R \in \{48, 64, 80, 96, 112\}$, covering a spectrum of visual fidelity with respect to the original standard resolution of $R = 64$ (Cobbe et al., 2020). We do not use frame stacking or grayscale conversion, adhering to the standard Procgen protocol.

Figure 2 visualizes example images from four Procgen-HD environments, with the remaining 12 games detailed in Appendix A.1. Lower-resolution renderings exhibit severe visual artifacts that obscure crucial details, e.g., determining the orientation of the green ego fish in *Bigfish* or the ego aircraft in *Starpilot* becomes challenging. Crucially, increasing the resolution in Procgen-HD does *not* expand the field of view, which remains constant, but solely increases the information density of the visual observation.

Network Architecture: For image processing, all experiments use an identical ResNet-based backbone derived from the Impala architecture (Espeholt et al., 2018), which has been widely adopted as the standard in recent state-of-the-art approaches. As visualized in Figure 3, this backbone processes the input observation x through a stack of three hierarchical convolutional sequences `ConvSeq`. Each sequence consists of a Conv2D layer with stride=1, followed by a max-pooling operation with stride=2 and a series of two residual blocks

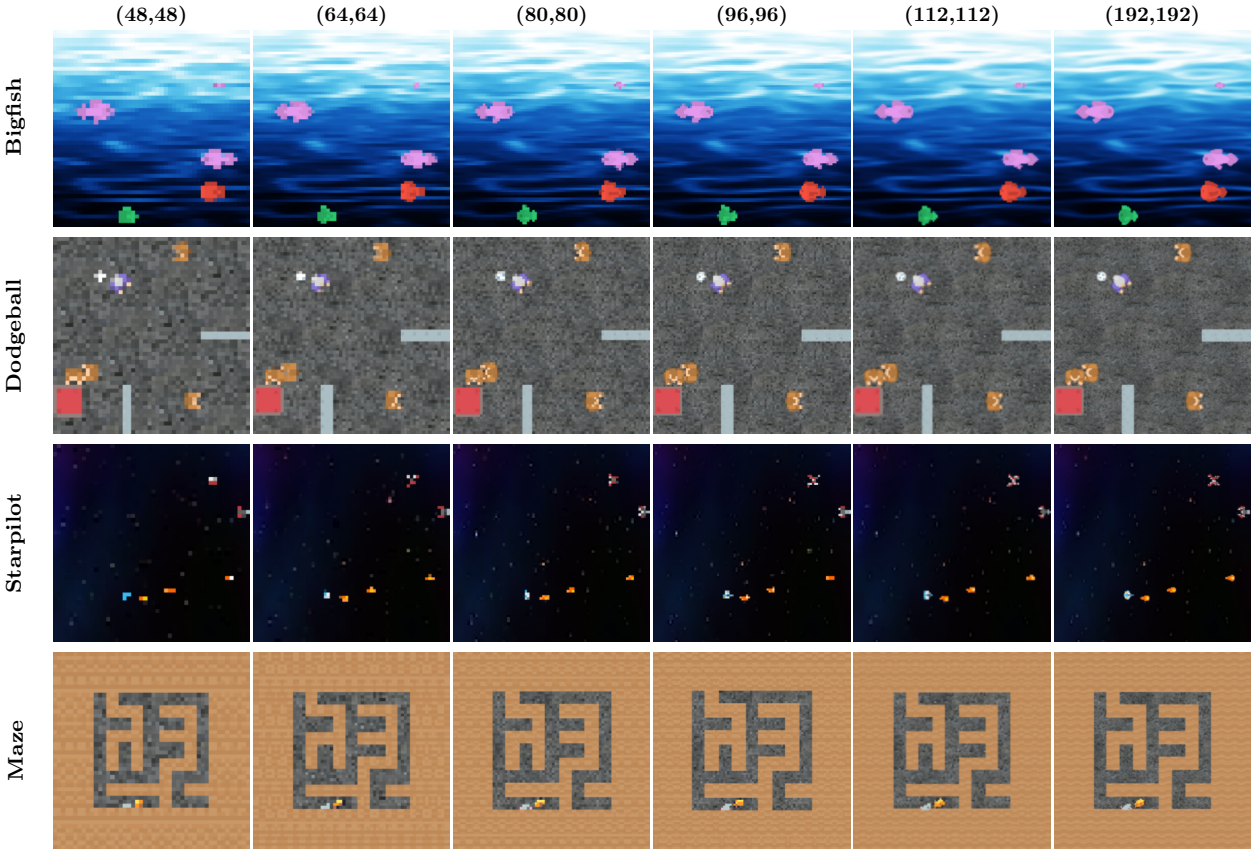


Figure 2: Comparison of a subset of 4 Procgen-HD environments at different image resolutions of (R, R) pixels. All images depict the same scene, rendered at varying resolutions $R \in \{48, 64, 80, 96, 112, 192\}$ with the field of view remaining constant. The $(192, 192)$ resolution is provided solely for visual comparison against an image with minimal compression artifacts but not further evaluated.

`ResBlock`. The `ResBlock` units consist of two Conv2d layers with stride=1. For all Conv2d layers in each of the three `ConvSeq`, the standard configuration uses the same number of $\{16, 32, 32\}$ filters, respectively.

Following prior work on scaling the Impala network (Cobbe et al., 2020; Trumpp et al., 2025), we introduce a width multiplier τ. This parameter scales the number of filters per convolutional layer, allowing us to vary the representational capacity while keeping the backbone structure constant. Based on the results of Trumpp et al. (2025), we only examine already scaled networks with $\tau \in \{2, 3, 4\}$.

We test both the original version of Impala and the recently proposed derivative Impoola:

- **Impala (Standard/Flatten)** (Espeholt et al., 2018): The standard approach flattens the final $H \times W \times C$ feature map into a single vector. Crucially, this vector's size grows quadratically with input resolution. This causes a massive increase in the parameter count of the first dense layer, directly linking model size to visual fidelity as visualized in Figure 3.
- **Impoola (GAP):** As proposed by Trumpp et al. (2025), this variant applies global average pooling (GAP) after the final residual block, similar to Sokar & Castro (2025). GAP aggregates spatial dimensions into a fixed-size vector of dimension C. This effectively decouples the parameter count from the input resolution, ensuring the model size remains *constant* even as visual fidelity increases.

Training Settings: Unless otherwise stated, our results encompass the *full* benchmark across all 16 environments. Following the established standard (Cobbe et al., 2020), we train for 25M timesteps in the *easy* setting and 100M timesteps in the *hard* setting.

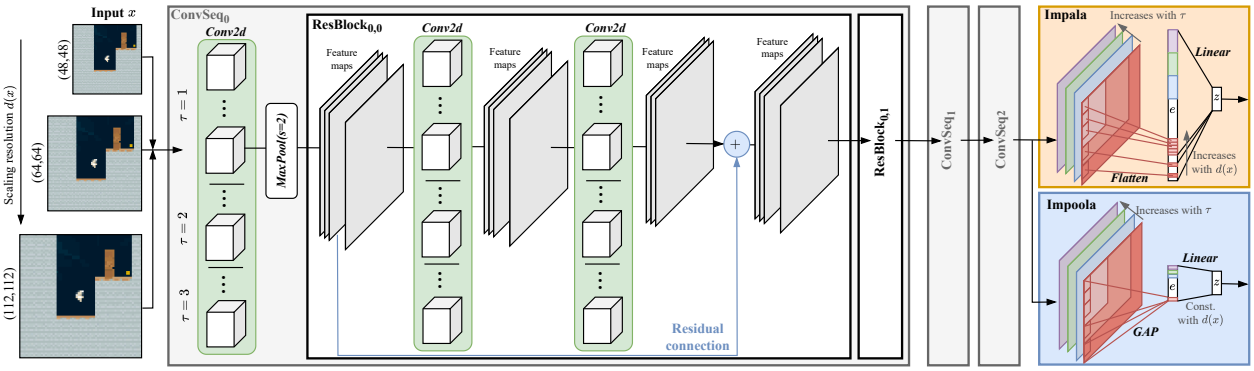


Figure 3: The standard flattening approach in Impala (**top**) leads to parameter explosion as image resolution d(x) increases. In contrast, the Impoola architecture (**bottom**) utilizes global average pooling (GAP) to maintain a constant latent representation size when increasing the input image resolution. The parameter τ is a scaling factor defining the number of filters in each Conv2d layer. Figure derived with permission from Trumpp et al. (2025).

We primarily focus on the generalization track, which allows us to measure generalization by defining distinct sets of levels for training and testing. The *easy* setting restricts training to 200 simpler levels, whereas the *hard* setting uses 1000 levels with significantly increased game difficulty and complexity. The additional results for the *efficiency* track do not restrict the number of training levels.

DRL Algorithm: As proximal policy optimization (PPO) (Schulman et al., 2017) is the established baseline for Procgen, maintaining this standard allows us to isolate the impact of input resolution and architectural scaling without the confounding influence of algorithmic choice. The actor and critic for PPO share the image encoder; further implementation details and hyperparameters are listed in Appendix A.2.

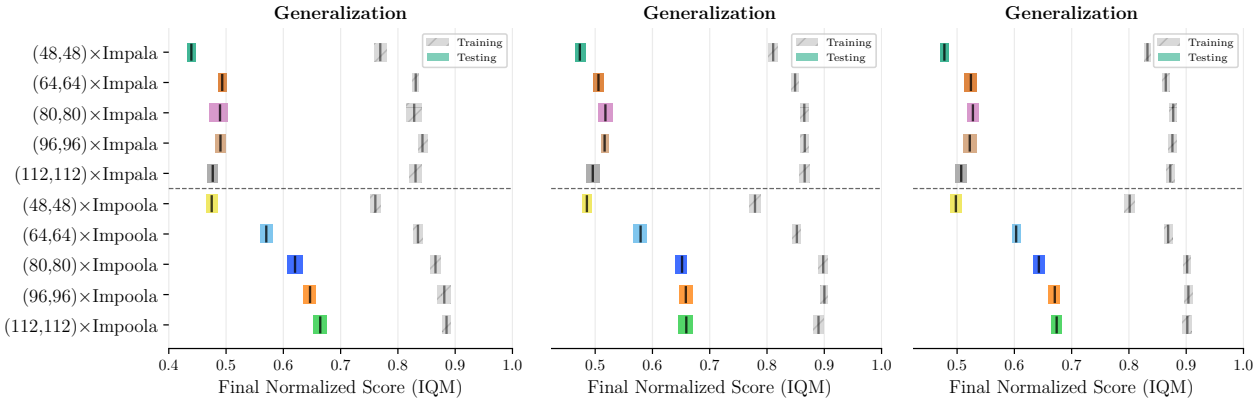
Evaluation Metrics: To measure generalization capability in the generalization track, we evaluate performance on the held-out test levels that the agent has never encountered during training. The efficiency track is evaluated on all levels. During training, we periodically collect episodic returns over 2,500 episodes using these unique levels. These returns are normalized to obtain scores using the constants from Cobbe et al. (2020), where 1.0 denotes optimal performance and 0.0 corresponds to a random policy. Each environment is evaluated with 5 independent runs using different random seeds. Aggregated results across environments are reported as interquantile mean (IQM) (Agarwal et al., 2021) scores, with 95% stratified bootstrap confidence intervals shown as shaded regions.

4 Empirical Results

We argue that observation resolution is a first-order variable that has been held fixed by convention rather than by design, and that this has left both architectural limitations and potential gains unexplored. We test this hypothesis empirically, asking whether increasing resolution improves performance and which architectures can leverage it (**RQ1**), whether these gains reflect genuine generalization (**RQ2**), whether the findings hold across training regimes (**RQ3**), and which perceptual demands drive them (**RQ4**).

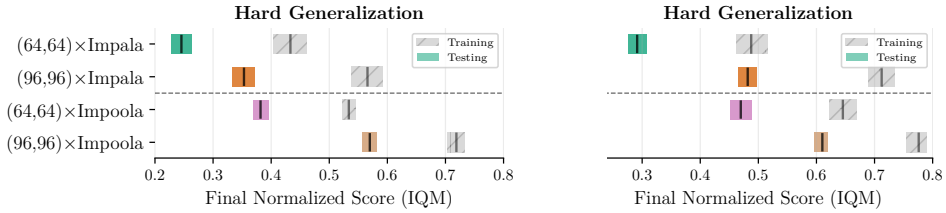
4.1 RQ1: Does Higher Resolution Improve Performance?

As shown in Figure 1, the effect of increasing resolution depends strongly on architecture. Impala shows a modest improvement from $(64, 64)$ to $(80, 80)$, but this trend does not persist and performance declines at higher resolutions across width scales, suggesting that additional visual detail is not effectively utilized. Notably, this improvement emerges only at $\tau = 3$, whereas lower capacity Impala variants do not benefit from increased resolution. In contrast, Impoola improves consistently with resolution across all capacity configurations. The best aggregate testing score of 0.674 is achieved by Impoola with $\tau = 4$ at $(112, 112)$, representing an approximate 18 % improvement only through scaling in comparison to itself at the standard



(a) Results for $\tau = 2$. (b) Results for $\tau = 3$. (c) Results for $\tau = 4$.

Figure 4: Comparison of the final aggregate results after 25 M training steps for the *easy* generalization across width scale $\tau \in \{2, 3, 4\}$ and image resolutions of $(48, 48)$ to $(112, 112)$. Evaluation of the training levels (gray) is shown alongside performance on held-out test levels (colored).



(a) Results for $\tau = 2$. (b) Results for $\tau = 3$.

Figure 5: Comparison of the final aggregate results after 100 M training steps for the *hard* generalization across width scale $\tau \in \{2, 3\}$ and input image resolutions of $(64, 64)$ and $(96, 96)$. Evaluation of the training levels (gray) is shown alongside performance on held-out test levels (colored).

configuration with $(64, 64)$ and $\tau = 2$. Impala's best score of 0.527 is achieved with $\tau = 4$ at $(80, 80)$, benefiting mostly from the increased network size with $\tau = 4$ rather than the higher image resolution. Ultimately, under their respective best conditions, Impoola outperforms Impala by 28 %.

This performance difference most likely arises from how each architecture handles the growth of spatial feature maps. As resolution increases, Impala produces larger feature maps whose flattened representation expands rapidly, limiting its ability to effectively exploit the additional spatial information. Impoola, by contrast, remains largely size-independent through global average pooling and therefore adapts seamlessly to higher resolutions, as previously summarized by Figure 1.

A natural question is whether increasing network depth can alleviate Impala's limitation. Adding a fourth `ConvSeq` block increases the model depth to 20 layers, thereby reducing the spatial size of the feature maps before the policy head and limiting the growth of the flattened representation. As shown in Figure 6, the deeper Impala model is able to benefit from increased resolution, confirming that the bottleneck lies in the flattening operation rather than in higher resolution itself. Nevertheless, Impoola still achieves larger gains with substantially fewer parameters. While prior Procgen scaling work (Cobbe et al., 2020) has primarily emphasized width, i.e., increasing the number of filters per Conv2d layer, these results suggest that depth provides a complementary axis for enabling effective resolution scaling.

4.2 RQ2: Does Higher Resolution Improve Generalization?

While RQ1 showed that higher resolution improves aggregate performance, we next ask whether these gains reflect genuine generalization or simply stronger fitting to the training levels.

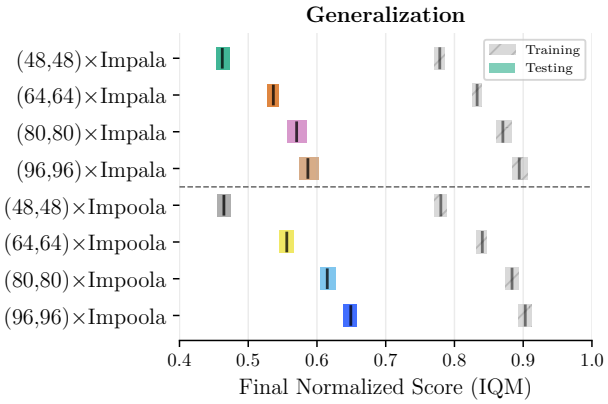


Figure 6: Aggregate results of testing levels in the generalization track for *deeper* networks, which have a fourth `ConvSeq` block with 32 filters in the unscaled setting added, totaling 20 layers. Results are shown for a width scale of $\tau = 3$.

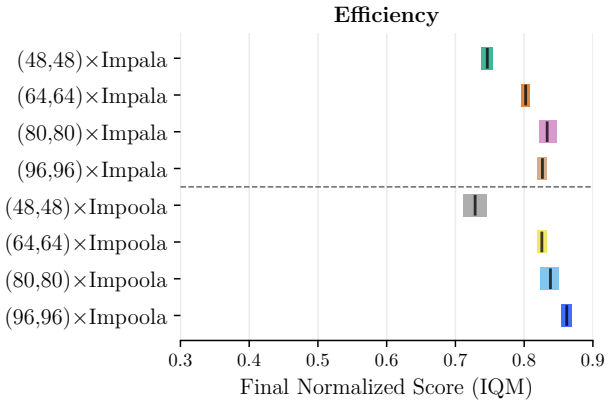


Figure 7: Aggregate testing results for the efficiency track with a width scale of $\tau = 3$. The efficiency track removes the level restriction entirely, exposing agents to the full procedural distribution during training.

In the easy generalization setting, where training is restricted to 200 levels, agents achieve near-optimal training scores of approximately 0.9 on these levels across most configurations. This training saturation compresses the visible benefit of higher resolution in the aggregated results, as improvements can only manifest on unseen test levels. To isolate the effect of resolution on generalization, we examine training and test performance separately.

Figure 4 shows that the relationship between resolution and generalization differs across architectures. For Impala, increasing resolution yields only small improvements in training performance, while test performance declines at $\tau = 2$ and improves only slightly at $\tau = 3$, leaving a persistent gap between training and test results. In contrast, Impoola exhibits a different pattern. Training performance changes only modestly with increasing resolution, whereas test performance improves substantially, narrowing the gap between them. This indicates that even small improvements during training translate into stronger gains on unseen levels. For $\tau = 4$, Impoola's train–test gap at $(48, 48)$ is approximately 0.3, which reduces to 0.23 at $(112, 112)$, consistent with improved generalization at higher resolution. Even within this saturating regime, Impoola's consistent gains confirm that the standard $(64, 64)$ rendering discards visual detail that, when preserved, measurably improves generalization. Furthermore, as shown in Appendix Figure 14, a recent tokenization-based architecture (Sokar et al., 2025) similarly exhibits substantial gains from resolution scaling, yet it remains outperformed by Impoola.

4.3 RQ3: How Robust Are These Findings Across Training Regimes?

To determine how the impact of resolution depends on the training regime, we evaluate two additional Procgen-HD tracks that vary the task difficulty and data diversity.

Hard Generalization Track. We evaluate $\tau = 2$ and $\tau = 3$ at $(64, 64)$ versus $(96, 96)$. As shown in Figure 5, increasing resolution leads to clear performance improvements for both architectures. With 1000 training levels and substantially harder tasks, agents operate further from optimal performance, allowing additional visual detail to translate more directly into improved decision quality. Increasing network width further strengthens these gains. Although Impala continues to exhibit a larger train–test gap than Impoola, it still benefits substantially from higher resolution, while Impoola consistently converts increased visual fidelity into stronger performance across both width scales. Overall, these results indicate that when sufficient headroom remains, higher resolution reliably improves performance across architectures.

Efficiency Track. The efficiency track removes the level restriction entirely, exposing agents to the full procedural distribution during training. As shown in Figure 7, the gains from increasing resolution are

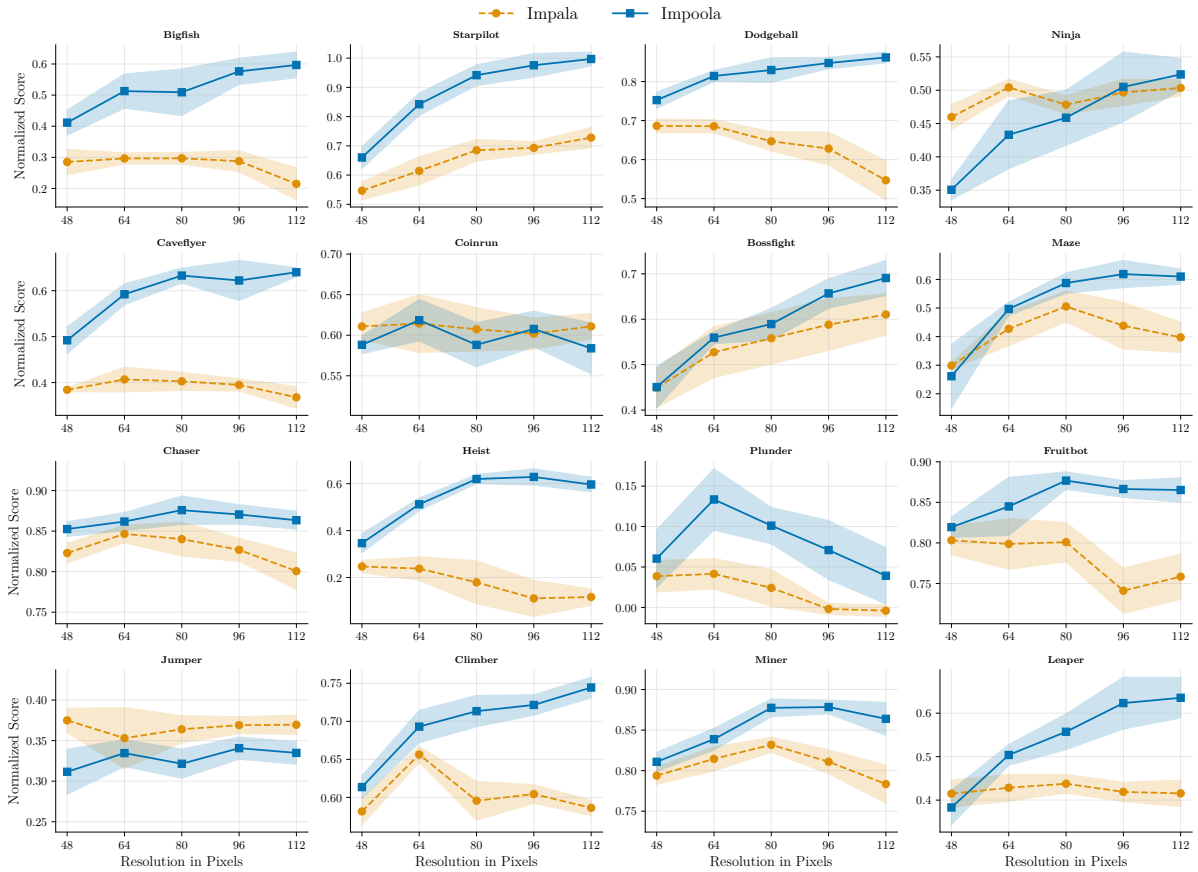


Figure 8: Environment-level comparison for all 16 Procgen-HD games for generalization, showing the final normalized scores of Impala and Impoola for testing levels. Both architectures use a width scale of $\tau = 4$; the resolution increases from $(48, 48)$ to $(112, 112)$.

more modest in this setting, which is expected: unlike the generalization track, i.e., training is restricted to 200 levels, agents now encounter the full level diversity during training, leaving less room for resolution to improve coverage of unseen variations. Nevertheless, Impoola continues to benefit from higher resolution and outperforms Impala on average across resolutions, confirming that the resolution gains observed in the generalization setting reflect genuine perceptual improvements rather than artifacts of limited training diversity. Reducing resolution below $(64, 64)$ degrades performance for both models, confirming that the standard resolution represents a meaningful threshold below which critical visual information is lost.

4.4 RQ4: Which Environments Benefit Most from Higher Resolution?

The aggregate results show that resolution scaling improves performance, but aggregate metrics alone do not reveal where these gains originate. To identify which environments benefit most from higher resolution, we analyze environment-level performance in Figure 8

The largest gains occur in environments requiring precise perception of small or distant entities. In *Dodgeball* and *Starpilot*, where the agent must accurately perceive and track relatively small, moving objects, increasing resolution from 64×64 to 112×112 yields substantial performance leaps. Similarly, in *Maze* and *Heist*, higher resolution likely improves visibility of distant structural cues and key objects, enabling more efficient navigation. In these environments, perceptual fidelity is a clear bottleneck at standard resolution.

In contrast, environments dominated by large static structures or simple locomotion, such as *Coinrun*, *Chaser*, and *Jumper*, show diminishing returns from resolution scaling. Because the standard $(64, 64)$ resolution already resolves critical obstacles, improvements remain modest, indicating that difficulty in these

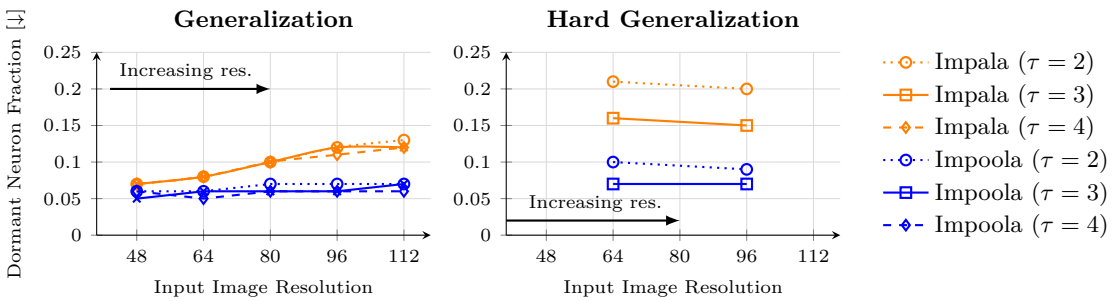


Figure 9: Comparison of dormant neurons for generalization with the easy (**left**) and hard (**right**) setting.

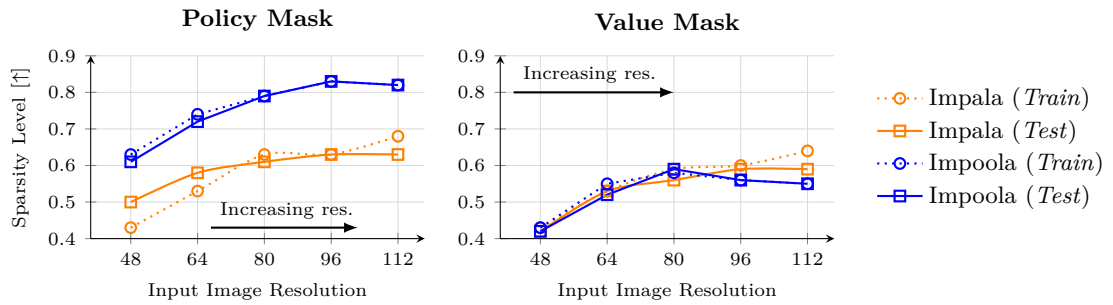


Figure 10: Comparison of policy (**left**) and value (**right**) mask sparsity, based on saliency maps and networks with a width scale of $\tau = 2$ for the generalization track. Saliency maps are calculated for observations obtained for the restricted testing and training levels, calculated as the mean over all observations encountered in 5 full evaluation episodes, and using a threshold of $\epsilon = 0.1$ for masking.

environments is driven more by dynamics than perception. It is noteworthy that *Jumper* and *Coinrun* are games with *agent-centered* observations, so these results overlap with the findings from Trumpp et al. (2025), who discuss that Impala often possesses an initial, advantageous inductive bias for such games in comparison to Impoola. However, as demonstrated by the trend in *Ninja*, our findings show that scaling the input resolution may eventually help Impoola outperform Impala for such games as well.

Additionally, while resolution scaling does not necessarily resolve general convergence issues, e.g., as found for Plunder, the trend towards improved performance from resolution scaling is significantly intensified in the *hard* generalization setting. As seen in Figure 19, the increased difficulty makes higher visual fidelity advantageous across *all* environments for Impoola, and it also benefits Impala in the majority of games.

5 Representational Analysis

We examine *why* Impoola benefits more from higher resolution than Impala by analyzing two properties of the learned networks: the fraction of dormant neurons (Sokar et al., 2023), which measures how broadly each architecture utilizes its capacity, and gradient-based saliency (Wang et al., 2016), which reveals where in the visual input the network directs its attention.

5.1 Dormant Neurons

In the easy generalization track shown in Figure 9 (left), Impala exhibits a clear monotonic increase in dormant neurons as resolution grows. Despite the network gaining additional capacity at higher resolutions, a progressively larger fraction of neurons remains inactive, indicating that this capacity is not effectively translated into richer representations. In contrast, Figure 9 (right) reveals a different pattern for the hard generalization track: dormant neuron fractions slightly decrease with resolution, suggesting that higher task

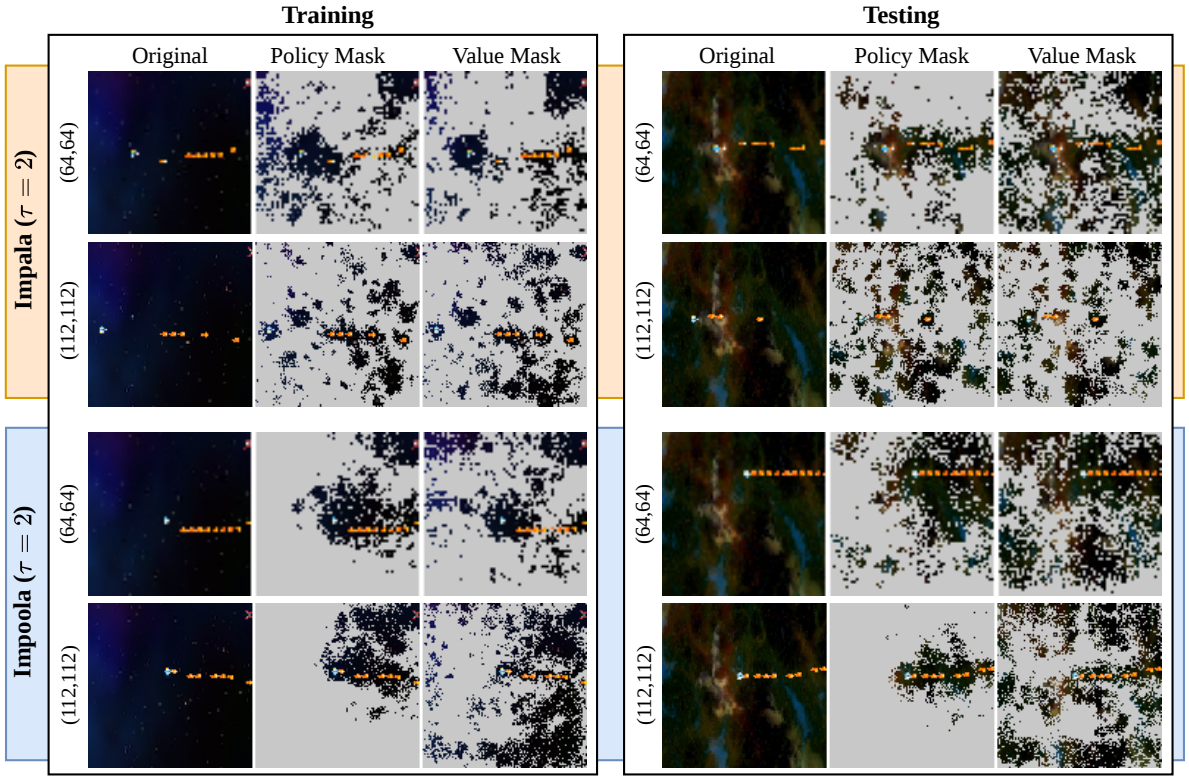


Figure 11: Qualitative comparison of saliency-based masks for the *Starpilot* game. Masks are binary overlays generated where the normalized saliency map exceeds a threshold of $\epsilon = 0.1$. The visualization highlights the specific regions retained by the Policy and Value heads for the Impala and Impoola agents, demonstrating the sparsity of the input processing at varying image resolutions. Pixels that are below the masking threshold are visualized as gray areas.

complexity encourages broader use of the network's capacity. Notably, these trends are consistent with the performance results in each setting, where resolution gains are larger in the hard track.

Across both settings, Impoola consistently maintains lower and more stable dormant neuron fractions, with the gap becoming particularly pronounced in the harder track. This suggests that GAP-based architectures better preserve active representations as resolution increases, rather than accumulating unused capacity.

5.2 Saliency Maps

The dormant neuron analysis characterizes how each architecture utilizes its internal capacity, but does not reveal *where* in the visual input the network directs its attention. To investigate this, we employ gradient-based saliency analysis, computing the gradient of the value function $\nabla_x V(x)$ and the policy logit $\nabla_x \log \pi(a \mid x)$ with respect to the input observation s, following saliency visualization methods used for deep RL agents (Wang et al., 2016). For each frame, we take the absolute gradient magnitude, normalize it to $[0, 1]$ using 99th-percentile clipping, and apply a binary threshold mask at $\epsilon = 0.1$: pixels whose normalized gradient exceeds ϵ are marked as attended, while the remainder are suppressed. We summarize these masks with a scalar *mask sparsity* metric, defined as the fraction of pixels below the threshold. Higher sparsity indicates that the network concentrates its gradient signal on a smaller, more localized region. We conduct the following analysis on *Starpilot*, where the agent must track small, fast-moving projectiles and enemies across diverse backgrounds, and where resolution scaling yields a substantial performance gain.

As shown in Figure 10 (right), the value mask sparsity remains relatively flat across resolutions for both architectures, and the corresponding masks in Figure 11 show broad activation covering large portions of the

scene. This is expected: $V(x)$ captures a global assessment of the state, so its gradient signal naturally spreads across the full observation rather than localizing on specific entities. The policy mask sparsity presented in Figure 10 (left) reveals a contrasting pattern. Impoola's sparsity increases steadily from approximately 0.72 at $(64, 64)$ to 0.82 at $(112, 112)$, meaning the network attends to a progressively smaller fraction of the input as resolution grows, concentrating tightly on the agent, nearby enemies, and projectiles while suppressing background regions; see Figure 11. Impala's policy sparsity, by contrast, remains largely constant on test levels, and its masks stay diffuse even at higher resolutions.

We reason that at low resolutions, small objects are obscured by blurring and aliasing, whereas higher resolution restores the visual distinctness required for precise localization. However, the fact that this increased sparsity appears in Impoola but not in Impala suggests that the architecture must be able to process the additional spatial detail effectively for it to translate into more focused policy attention.

6 Limitations

A central challenge in studying resolution scaling is the high computational cost of processing high-dimensional inputs. Increasing the observation resolution from $(48, 48)$ to $(112, 112)$ results in a substantial increase in floating-point operations and memory usage. We visualize this cost in Appendix Figure 21, e.g., increasing from Procgen's standard resolution of $(64, 64)$ to $(112, 112)$ results in an approximately $2.7\times$ longer total training time; Impoola and Impala have the same training times since the cost of the Linear layer is negligible compared to the Conv2d layers. As such, this intense computational demand necessitated a focused scope; our study's results are already based on more than 20,000 A100 GPU hours for training. We therefore ground our analysis in the Procgen benchmark, specifically adopting the PPO algorithm and the Impala and Impoola models in accordance with the benchmark's established standards.

Our choice of the Procgen benchmark offers particular advantages over traditional suites like Atari, as it explicitly evaluates generalization across a diverse set of games with varying visual characteristics, enabling robust assessment of an agent's perceptual capabilities. However, since our study is conducted entirely within this procedural, discrete-action domain, it should be seen as a first compelling demonstration of visual scaling laws at higher resolutions, but further exploration is required. As such, we hope the community will extend our effort by investigating continuous control tasks, 3D environments, and off-policy methods in the future.

7 Conclusion

This work challenges the prevailing assumption that low-resolution inputs are sufficient for deep RL. Using Procgen-HD, a configurable-resolution variant of the Procgen benchmark, we show that higher resolution substantially improves both performance and generalization, specifically when paired with a resolution-independent architecture. The standard Impala encoder stagnates at higher resolutions due to quadratic parameter growth in its flattening layer, while Impoola's global average pooling decouples capacity from resolution and achieves a 28 % improvement at their respective best conditions. Environment-level analysis reveals that these gains concentrate where precise perception of small or distant entities matters most, and gradient saliency analysis implies that a key mechanism is an improved spatial localization in the policy network. These findings suggest that the conventional $(64, 64)$ and $(84, 84)$ defaults are not architecturally neutral but instead impose a perceptual ceiling that favors resolution-dependent designs and limits scalability. We hope these initial findings encourage broader study of the impact of resolution scaling in visual deep RL. We also see particular relevance for the robotics community, which often deals with high-resolution images.

References

Zaheer Abbas, Rosie Zhao, Joseph Modayil, Adam White, and Marlos C Machado. Loss of plasticity in continual deep reinforcement learning. In *Conference on lifelong learning agents*, pp. 620–636. PMLR, 2023.

Rishabh Agarwal, Max Schwarzer, Pablo Samuel Castro, Aaron C Courville, and Marc Bellemare. Deep reinforcement learning at the edge of the statistical precipice. *Advances in neural information processing systems*, 34:29304–29320, 2021.

Marc G Bellemare, Yavar Naddaf, Joel Veness, and Michael Bowling. The arcade learning environment: An evaluation platform for general agents. *Journal of artificial intelligence research*, 47:253–279, 2013.

Alexander Braylan, Mark Hollenbeck, Elliot Meyerson, and Risto Miikkulainen. Frame skip is a powerful parameter for learning to play atari. In *AAAI workshop: Learning for general competency in video games*, 2015.

Johan Samir Obando Ceron, Ghada Sokar, Timon Willi, Clare Lyle, Jesse Farebrother, Jakob Nicolaus Foerster, Gintare Karolina Dziugaite, Doina Precup, and Pablo Samuel Castro. Mixtures of experts unlock parameter scaling for deep RL. In *Forty-first international conference on machine learning*, 2024.

Karl Cobbe, Chris Hesse, Jacob Hilton, and John Schulman. Leveraging procedural generation to benchmark reinforcement learning. In *International conference on machine learning*, pp. 2048–2056. PMLR, 2020.

Mostafa Dehghani, Yi Tay, Alexey A Gritsenko, Zhe Zhao, Neil Houlsby, Fernando Diaz, Donald Metzler, and Oriol Vinyals. The benchmark lottery. *arXiv preprint arXiv:2107.07002*, 2021.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, 2009.

Gabriel Dulac-Arnold, Daniel Mankowitz, and Todd Hester. Challenges of real-world reinforcement learning. *arXiv preprint arXiv:1904.12901*, 2019.

Lasse Espeholt, Hubert Soyer, Remi Munos, Karen Simonyan, Vlad Mnih, Tom Ward, Yotam Doron, Vlad Firoiu, Tim Harley, Iain Dunning, et al. Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures. In *International conference on machine learning*, pp. 1407–1416. PMLR, 2018.

Lasse Espeholt, Raphaël Marinier, Piotr Stanczyk, Ke Wang, and Marcin Michalski. Seed rl: Scalable and efficient deep-rl with accelerated central inference. In *International conference on learning representations*, 2020.

Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639): 115–118, 2017.

Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A Wichmann. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11):665–673, 2020.

Laura Graesser, Utku Evci, Erich Elsen, and Pablo Samuel Castro. The state of sparse training in deep reinforcement learning. In *International conference on machine learning*, pp. 7766–7792. PMLR, 2022.

Matthew J Hausknecht and Peter Stone. Deep recurrent q-learning for partially observable mdps. In *AAAI fall symposia*, volume 45, pp. 141, 2015.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019.

Julian Ibarz, Jie Tan, Chelsea Finn, Mrinal Kalakrishnan, Peter Pastor, and Sergey Levine. How to train your robot with deep reinforcement learning: lessons we have learned. *The International Journal of Robotics Research*, 40(4-5):698–721, 2021.

Steven Kapturowski, Georg Ostrovski, Will Dabney, John Quan, and Remi Munos. Recurrent experience replay in distributed reinforcement learning. In *International conference on learning representations*, 2019.

Bernard Koch, Emily Denton, Alex Hanna, and Jacob Gates Foster. Reduced, reused and recycled: The life of a dataset in machine learning research. In *Conference on neural information processing systems datasets and benchmarks Track (Round 2)*, 2021.

Jacob Eeuwe Kooi, Zhao Yang, and Vincent Francois-Lavet. Hadamax encoding: Elevating performance in model-free atari. In *Conference on neural information processing systems*, 2025.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images, 2009.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, volume 25. Curran Associates, Inc., 2012.

Aviral Kumar, Rishabh Agarwal, Dibya Ghosh, and Sergey Levine. Implicit under-parameterization inhibits data-efficient deep reinforcement learning. In *International conference on learning representations*, 2021.

Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(39):1–40, 2016.

Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2014.

Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *IEEE conference on computer vision and pattern recognition*, pp. 2117–2125, 2017.

Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986, 2022.

Clare Lyle, Mark Rowland, and Will Dabney. Understanding and preventing capacity loss in reinforcement learning. In *International conference on learning representations*, 2022.

Marlos C Machado, Marc G Bellemare, Erik Talvitie, Joel Veness, Matthew Hausknecht, and Michael Bowling. Revisiting the arcade learning environment: Evaluation protocols and open problems for general agents. *Journal of Artificial Intelligence Research*, 61:523–562, 2018.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin A. Riedmiller, Andreas Kirkeby Fidjeland, Georg Ostrovski, Stig Petersen, Charlie Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518:529–533, 2015.

Evgenii Nikishin, Max Schwarzer, Pierluca D'Oro, Pierre-Luc Bacon, and Aaron Courville. The primacy bias in deep reinforcement learning. In *International conference on machine learning*, pp. 16828–16847. PMLR, 2022.

Otávio A. B. Penatti, Keiller Nogueira, and Jefersson A. dos Santos. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In *IEEE conference on computer vision and pattern recognition workshops*, pp. 44–51, 2015.

Inioluwa Deborah Raji, Emily Denton, Emily M. Bender, Alex Hanna, and Amandalynne Paullada. AI and the everything in the whole wide world benchmark. In *Conference on neural information processing systems datasets and benchmarks Track (Round 2)*, 2021.

Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Ghada Sokar and Pablo Samuel Castro. Mind the GAP! the challenges of scale in pixel-based deep reinforcement learning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Ghada Sokar, Elena Mocanu, Decebal Constantin Mocanu, Mykola Pechenizkiy, and Peter Stone. Dynamic sparse training for deep reinforcement learning. In *International joint conference on artificial intelligence*, pp. 3437–3443, 7 2022.

Ghada Sokar, Rishabh Agarwal, Pablo Samuel Castro, and Utku Evci. The dormant neuron phenomenon in deep reinforcement learning. In *International conference on machine learning*, pp. 32145–32168. PMLR, 2023.

Ghada Sokar, Johan Samir Obando Ceron, Aaron Courville, Hugo Larochelle, and Pablo Samuel Castro. Don't flatten, tokenize! unlocking the key to softmoe's efficacy in deep RL. In *International conference on learning representations*, 2025.

Xingyou Song, Yiding Jiang, Stephen Tu, Yilun Du, and Behnam Neyshabur. Observational overfitting in reinforcement learning. In *International conference on learning representations*, 2020.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pp. 6105–6114. PMLR, 2019.

Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pp. 10096–10106. PMLR, 2021.

Hugo Touvron, Andrea Vedaldi, Matthijs Douze, and Hervé Jégou. Fixing the train-test resolution discrepancy. *Advances in neural information processing systems*, 32, 2019.

Raphael Trumpp, Ansgar Schäfftlein, Mirco Theile, and Marco Caccamo. Impoola: The power of average pooling for image-based deep reinforcement learning. *Reinforcement Learning Journal*, 6:1025–1047, 2025.

Saran Tunyasuvunakool, Alistair Muldal, Yotam Doron, Siqi Liu, Steven Bohez, Josh Merel, Tom Erez, Timothy Lillicrap, Nicolas Heess, and Yuval Tassa. dm_control: Software and tasks for continuous control. *Software Impacts*, 6:100022, 2020.

Ziyu Wang, Tom Schaul, Matteo Hessel, Hado Hasselt, Marc Lanctot, and Nando Freitas. Dueling network architectures for deep reinforcement learning. In *International conference on machine learning*, pp. 1995–2003. PMLR, 2016.

Steven D Whitehead and Dana H Ballard. Learning to perceive and act by trial and error. *Machine Learning*, 7(1):45–83, 1991.

Timon Willi, Johan Samir Obando Ceron, Jakob Nicolaus Foerster, Gintare Karolina Dziugaite, and Pablo Samuel Castro. Mixture of experts in a mixture of RL settings. *Reinforcement Learning Journal*, 3:1072–1105, 2024.

Denis Yarats, Ilya Kostrikov, and Rob Fergus. Image augmentation is all you need: Regularizing deep reinforcement learning from pixels. In *International conference on learning representations*, 2020.

Denis Yarats, Amy Zhang, Ilya Kostrikov, Brandon Amos, Joelle Pineau, and Rob Fergus. Improving sample efficiency in model-free reinforcement learning from images. In *AAAI conference on artificial intelligence*, volume 35, pp. 10674–10681, 2021.

Chiyuan Zhang, Oriol Vinyals, Remi Munos, and Samy Bengio. A study on overfitting in deep reinforcement learning. *arXiv preprint arXiv:1804.06893*, 2018.

A Experiment Details

A.1 Procgen-HD

Procgen-HD is a custom extension of the Procen Benchmark which allows for native rendering at different resolutions. There are the same 16 games available in Procgen-HD: Figure 2 shows *Bigfish*, *Dodgeball*, *Starpilot*, and *Maze*. The remaining 12 games are visualized in Figures 12 and Figure 13, respectively.

A.1.1 Observation Renderings

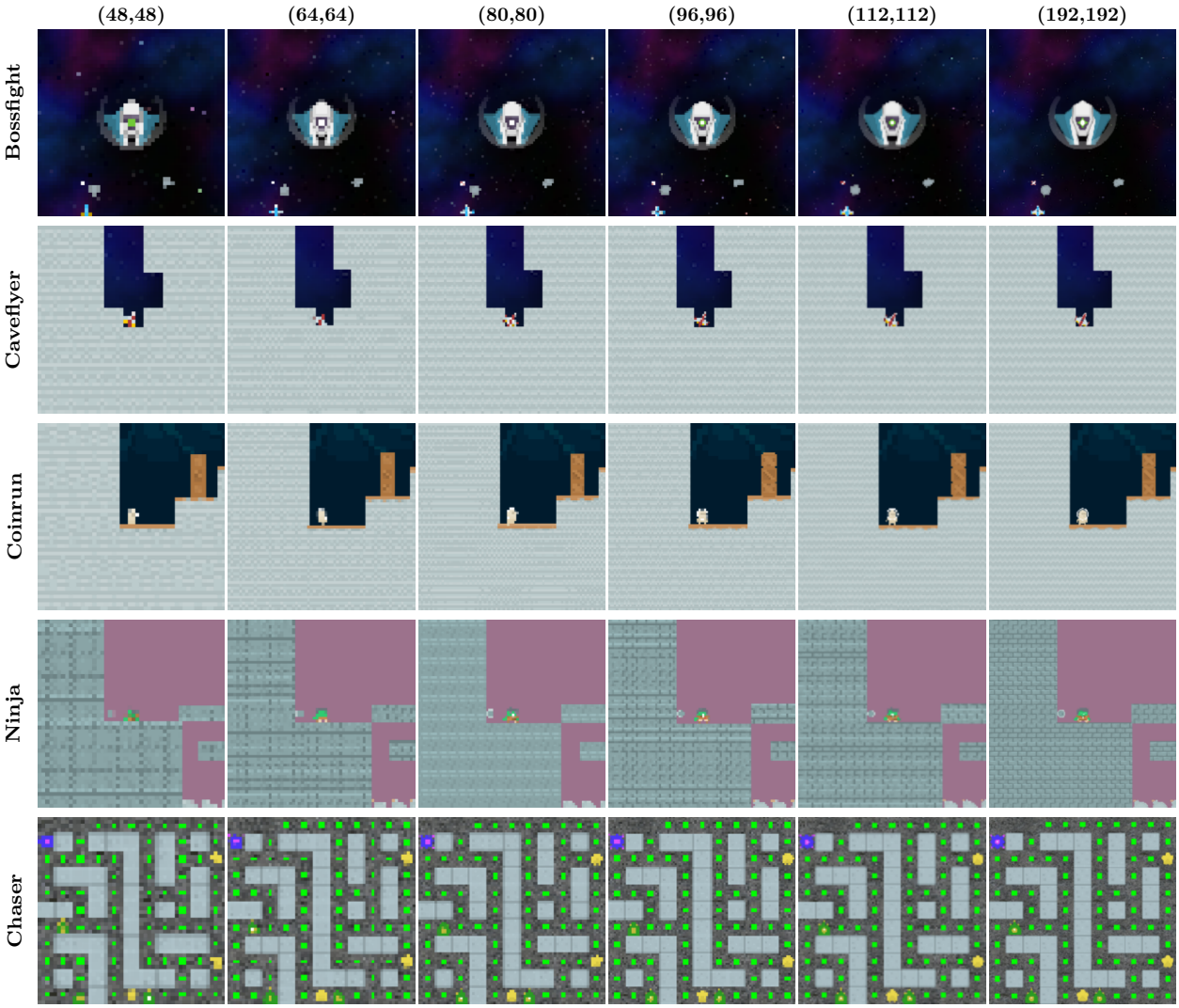


Figure 12: Comparison of further Procgen-HD environments at different image resolutions of (R, R) pixels. All images depict the same scene, rendered at varying resolutions $R \in \{48, 64, 80, 96, 112, 192\}$ with the field of view remaining constant. The $(192, 192)$ resolution is provided solely for visual comparison against an image with minimal compression artifacts but not further evaluated.

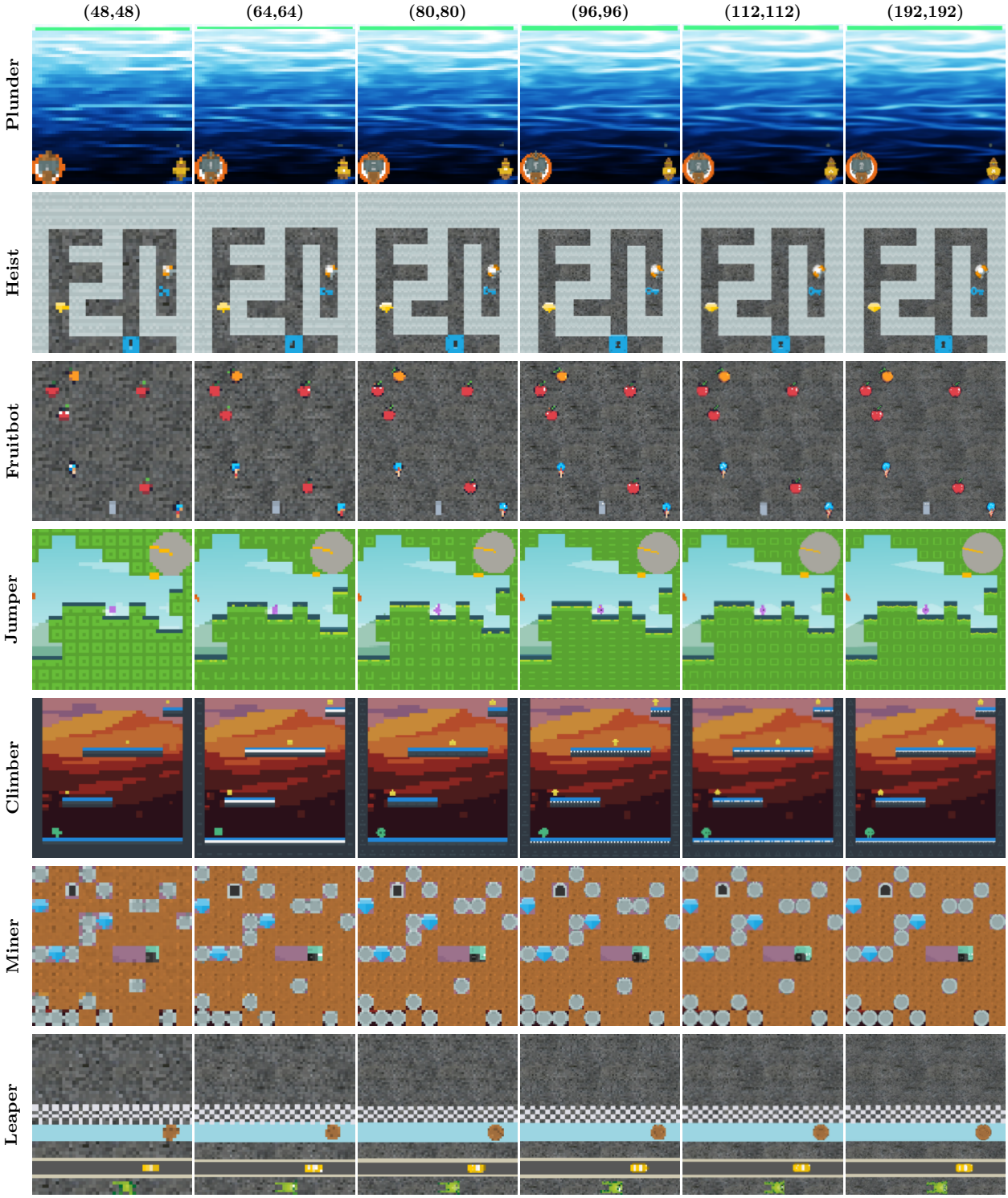


Figure 13: Comparison of further Procgen-HD environments at different image resolutions of (R, R) pixels. All images depict the same scene, rendered at varying resolutions $R \in \{48, 64, 80, 96, 112, 192\}$ with the field of view remaining constant. The $(192, 192)$ resolution is provided solely for visual comparison against an image with minimal compression artifacts but not further evaluated.

A.1.2 Game Characteristics

As discussed by Trumpp et al. (2025), there are four Procgen environments (*Coinrun*, *Jumper*, *Ninja*, and *Caveflyer*) where the map in the background is not fixed but translated in x- and y-direction relatively with the agent, i.e., the agent remains in the center of the observation, but may look left/right or rotate.

A.2 Hyperparameters List

Table 1: Hyperparameters for Proximal Policy Optimization (PPO).

Hyperparameter	**Values**
Number Parallel Environments	64
Environment Steps	256
Learning Rate ($\tau = 2$)	3.5×10^{-4}
Batch Size	2048
Epochs	3
Discount Factor γ	0.99
GAE Lambda (λ)	0.95
Clip Range	0.2
Value Function Coefficient	0.5
Entropy Coefficient	0.01
Max Gradient Norm	0.5
Optimizer	Adam
Shared Policy and Value Network	Yes

B Extended Empirical Results

B.1 Further Baseline

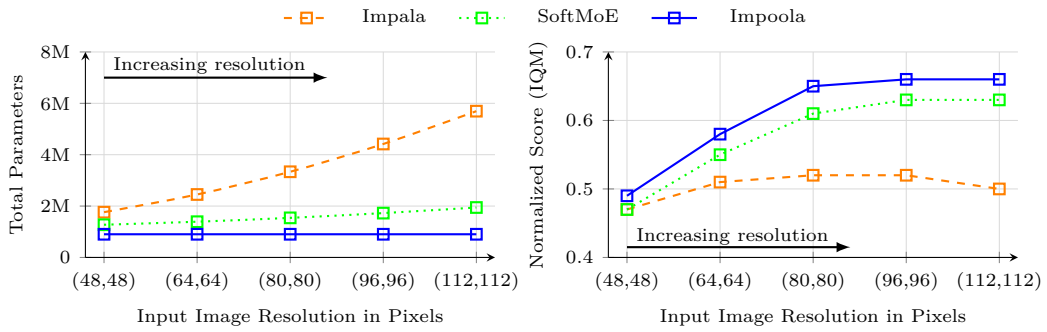


Figure 14: Environment-level comparison for all 16 Procgen-HD games for *generalization*, showing the final normalized scores for testing levels of Impala, Impoola, and additionally SoftMoE. All architectures use a width scale of $\tau = 3$; the resolution increases from $(64, 64)$ to $(112, 112)$. The counts include the parameters for separate actor and critic heads.

We evaluate the use of soft mixture-of-experts (SoftMoE) as proposed by Sokar et al. (2025) in comparison to the Impala and Impoola models. This soft mixture-of-experts (SoftMoE) uses the same backend as Impala, but instead of flattening the feature maps, they are tokenized and distributed to expert heads. Our results are based on the repository from Trumpp et al. (2025) (`https://github.com/raphajaner/impoola`), a PyTorch reimplementation of the official code. We use 10 expert heads and follow the standard parameters with the PerConv tokenization otherwise. As shown in Figure 14, SoftMoE scales to higher resolutions substantially better than the standard Impala baseline. However, SoftMoE's routing and slotting mechanism causes its

parameter count to still increase with resolution; this growth is less pronounced than Impala's rapid scaling, though. Ultimately, Impoola not only achieves superior overall performance, but it does so with a strictly resolution-independent parameter count and a significantly simpler network architecture and corresponding code implementation. This result supports our discussion that architectures that handle spatial features in a structured way can leverage additional visual detail, whereas Impala's flattening results in inefficient feature processing. As such, these results consistently extend the findings in Trumpp et al. (2025); Sokar & Castro (2025) that show the advantageous use of GAP, but also the potential of SoftMoEs.

B.2 Per Environment Resolution Scaling

B.2.1 Generalization

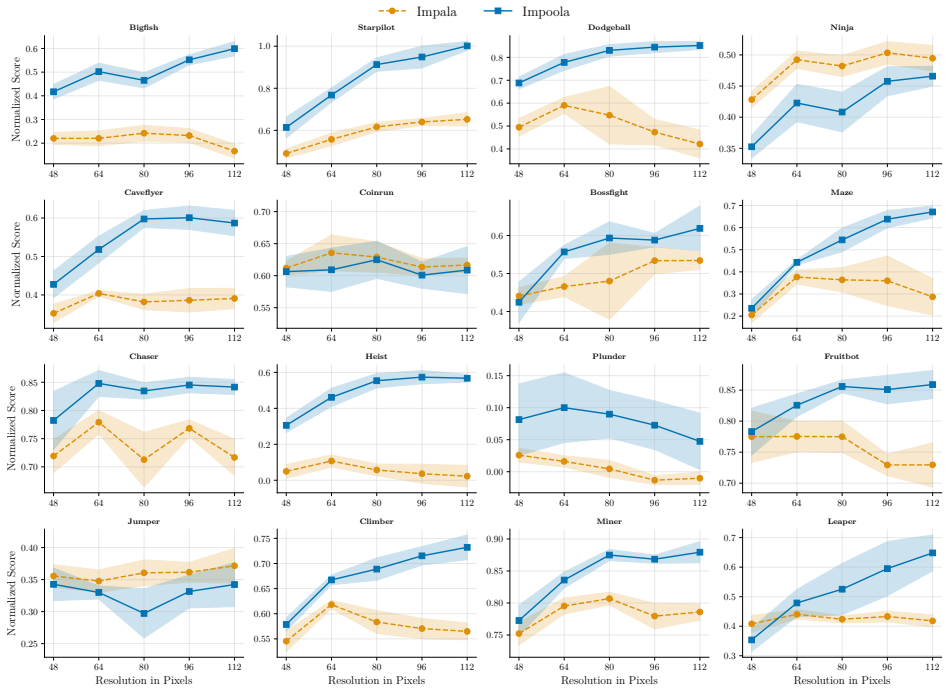


Figure 15: Environment-level comparison for all 16 Procgen-HD games for generalization, showing the final normalized scores of Impala and Impoola for testing levels. Both architectures use a width scale of $\tau = 2$; the resolution increases from $(48, 48)$ to $(112, 112)$.

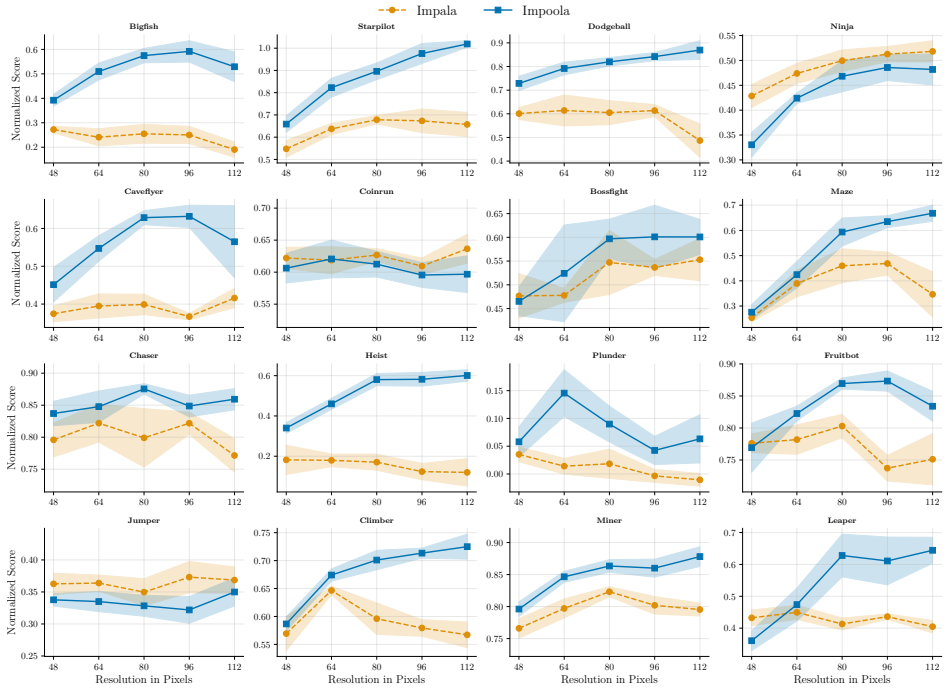


Figure 16: Environment-level comparison for all 16 Procgen-HD games for generalization, showing the final normalized scores of Impala and Impoola for testing levels. Both architectures use a width scale of $\tau = 3$; the resolution increases from $(48, 48)$ to $(112, 112)$.

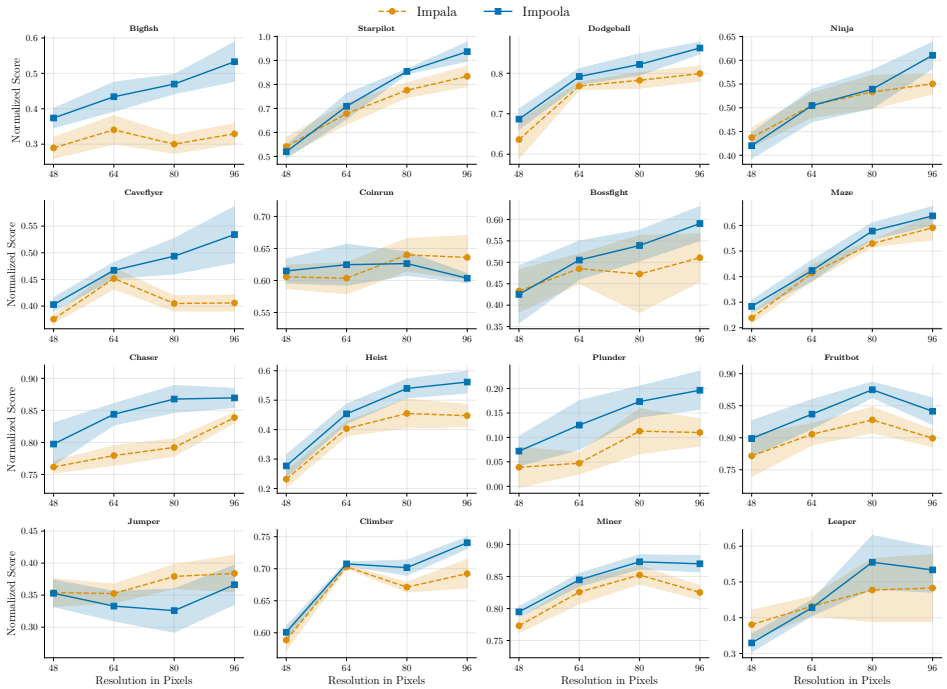


Figure 17: Environment-level comparison for all 16 Procgen-HD games for generalization, showing the final normalized scores of Impala and Impoola for testing levels. Both architectures use a *deeper* architecture with four `ConvSeq` elements and a width scale of $\tau = 3$; the resolution increases from $(48, 48)$ to $(112, 112)$.

B.2.2 Hard Generalization

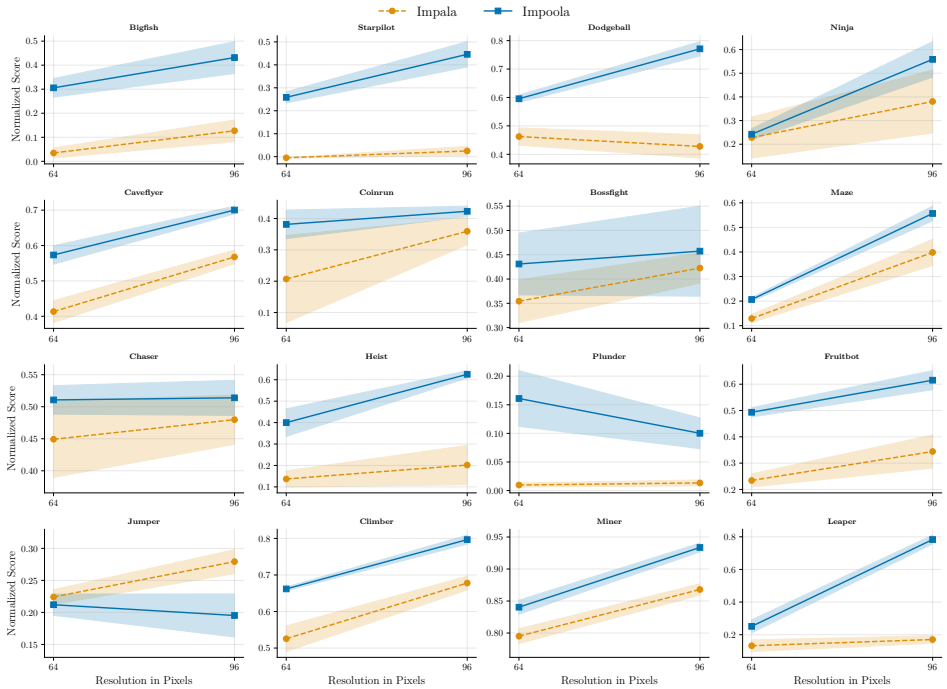


Figure 18: Environment-level comparison for all 16 Procgen-HD games for *hard* generalization, showing the final normalized scores of Impala and Impoola for testing levels. Both architectures use a width scale of $\tau = 2$; the resolution increases from $(64, 64)$ to $(96, 96)$.

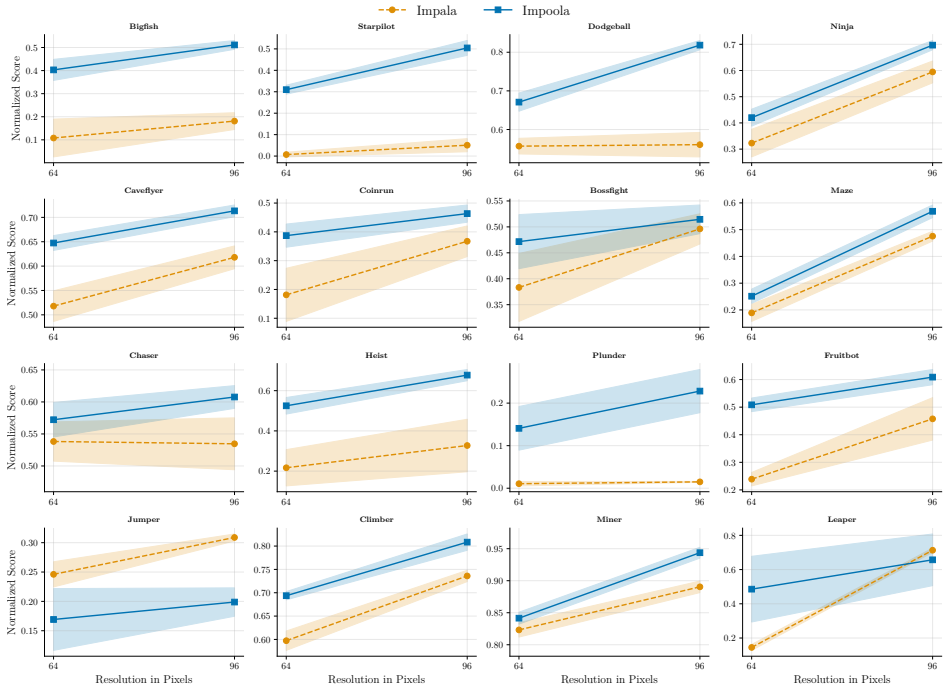


Figure 19: Environment-level comparison for all 16 Procgen-HD games for *hard* generalization, showing the final normalized scores of Impala and Impoola for testing levels. Both architectures use a width scale of $\tau = 3$; the resolution increases from $(64, 64)$ to $(96, 96)$.

B.2.3 Efficiency

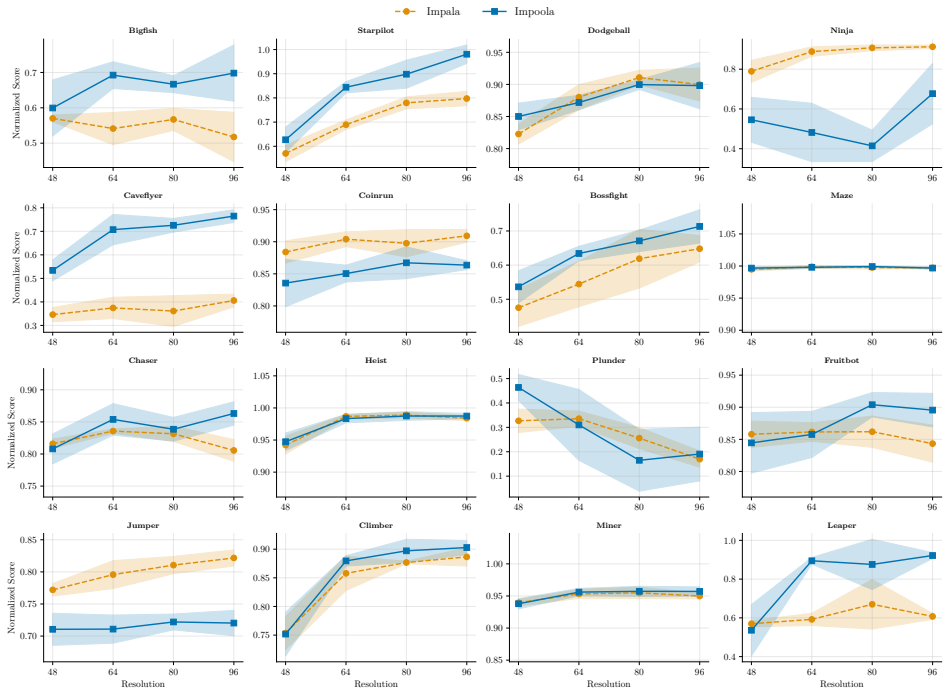


Figure 20: Environment-level comparison for all 16 Procgen-HD games for *efficiency*, showing the final normalized scores of Impala and Impoola for testing levels. Both architectures use a width scale of $\tau = 3$; the resolution increases from $(48, 48)$ to $(112, 112)$.

B.3 Computational Cost

Note: Presented times are approximate measurements using Python timers, averaged over 3 seeds for the Starpilot game, on a server with an AMD EPYC 7763 64-Core Processor (2P) CPU and an NVIDIA A100 PCIe 40GB GPU. Our code is based on PyTorch, but no `torch.compile()` is used. In each iteration, the *full* batch of recorded trajectories is loaded into VRAM, and mini-batch sizes according to Table A.2 are used. We use `torch.cuda.max_memory_allocated()` to measure VRAM usage.

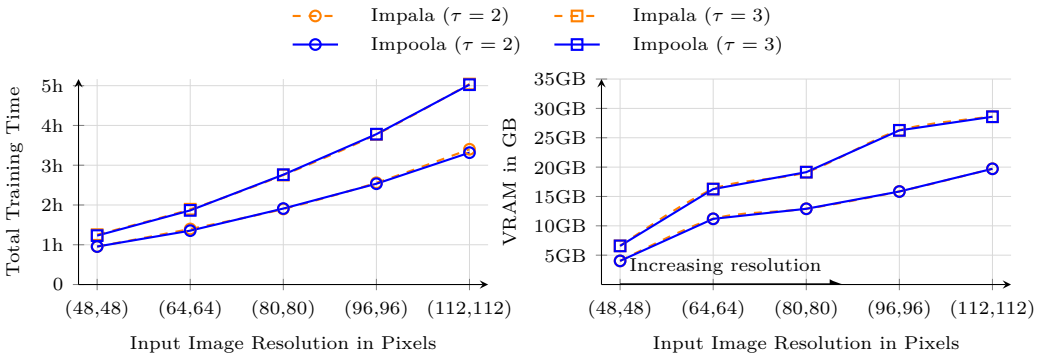


Figure 21: The impact of scaling the input image resolution on the total training time. We compare the common image encoders Impala and Impoola across resolutions from $(48, 48)$ to $(112, 112)$ pixels. Different network widths are shown, i.e., the number of filters per Conv2d layer is scaled by τ.

Figure 21 illustrates the impact of input resolution and network width scale τ on total training time and VRAM allocation. We find that Impoola and Impala exhibit almost identical computational costs across all tested configurations. Replacing the standard flattening operation with GAP does not provide a measurable benefit in either total training time or memory footprint. As such, the computational demand is driven entirely by the shared convolutional backbone and scales rapidly with the input resolution. For instance, increasing the resolution from $(64, 64)$ to $(112, 112)$ for a wider network with $\tau = 3$ inflates the training time from less than 2 h to approximately 5h, while VRAM consumption grows from roughly 16 GB to nearly 30 GB. We find the VRAM increase to be non-monotonic, e.g., exhibiting a plateau at the $(80, 80)$ resolution. We reason that this step-wise allocation behavior likely stems from dynamic hardware-level heuristics, such as cuDNN algorithm selection and cache size adjustments. In summary, it can be seen that the performance and generalization improvements from increased image resolutions can only be achieved when sufficient compute is available.

Furthermore, we also present which parts of the training contribute the most to the increase, splitting the training time for Impala and Impoola with $\tau = 3$ into 5 distinct steps as shown in Figure 22. It can be seen that the backward step for gradient calculation grows strongly with higher resolutions; thus, we see significant potential for speed improvements by optimizing the PyTorch implementation in future work.

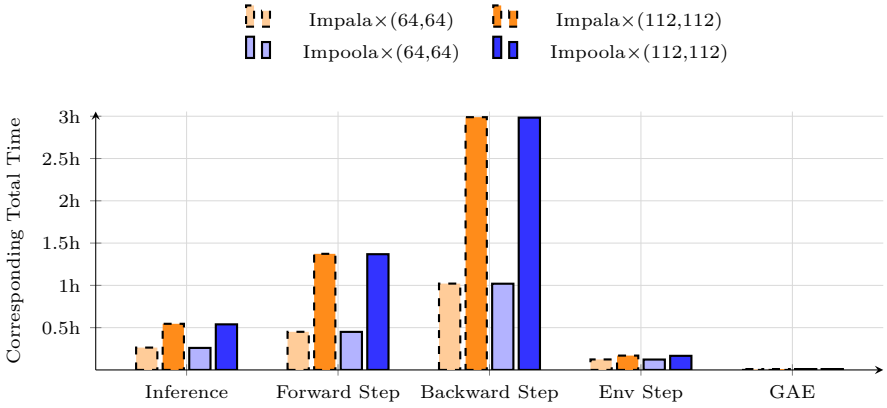


Figure 22: Breakdown of total training time, comparing the duration of the inference step (`a=model(x)`), environment step (`x'=env(a)`), GAE calculation, forward step (**loss=...**), and the backward step (`loss.backward()` calculation) for Impoola and Impala at $(64, 64)$ and $(112, 112)$ using $\tau = 3$. Other small steps included in the total training time, e.g., logging, is not depicted.