修士論文

注視点を導入した強化学習エージェントによる，視野の制約を前提とした行動獲得

勝又 翼

主指導教員 池田 心

北陸先端科学技術大学院大学
先端科学技術専攻
（情報科学）

令和７年３月

Abstract

In recent years, artificial intelligence (AI) technology has made remarkable progress, with research advancing in various fields such as image recognition and natural language processing. Games have often been used as benchmarks for AI techniques, and research has primarily focused on improving AI strength. AI has not only surpassed professional human players in perfect-information board games such as Go and shogi but has also achieved victories in team-based real-time digital games against professional human players. These results demonstrate that one of the primary goals of game AI research–creating AI that is stronger than human players–has been achieved in many games. However, AI behaviors optimized solely for strength often deviate significantly from human-like behaviors, posing challenges when using such AI as opponents or teachers for human players. Consequently, research has been conducted to develop game AI that exhibits human-like behavior.

An approach to achieving human-like game AI involves reinforcement learning with agents that share common human constraints on movement and perception. This approach aims to develop AI behaviors that resemble human actions and do not rely on superhuman reaction speeds or highly precise inputs by training agents in environments that incorporate the recognition accuracy and response speed limitations inherent to humans. In this work, we focus on a cognitive and behavioral characteristic of humans that has received little attention so far: the tendency to act while shifting their limited field of clear vision to observe desired areas.

In this work, we aim to investigate the learning of behaviors under visual field constraints and propose an approach that equips reinforcement learning agents with dynamic gaze points. The key features of this approach are the restriction of visual information based on the distance from the gaze point and the dynamic control of that gaze point. In this framework, the reinforcement learning agent must learn in an environment where the range of accurately perceivable information is limited and where it must actively control this range. To validate this approach, we conducted experiments by introducing visual field constraints in two different game genres.

In the first environment, a Breakout game, we first confirmed that blurring the input images created a visual handicap for the reinforcement learning agent. We then conducted reinforcement learning in an environment with visual field constraints. Specifically, we used Gaussian blur (smoothing) as a blurring method and prepared three settings: no blur, weak blur where the ball remained visible, and strong blur where the ball became unrecognizable. Reinforcement learning was conducted three times for each setting, and the results showed that performance

declined as the blur intensity increased. This demonstrated that the blurring process effectively functioned as a visual handicap for the agent's input.

Next, we conducted reinforcement learning in an environment where the game screen was divided into multiple areas, each with different levels of blur intensity. The game screen was divided into a 7×7 grid, where the 3×3 area centered on the gaze area remained unblurred. The weak blur was applied to the surrounding 7×7 area, while the strong blur was applied beyond that, simulating visual field constraints. The gaze area could be moved up, down, left, and right, and the agent's actions were defined as a combination of this movement and in-game actions. We created the environment so that the agent had to learn both "how to act in the game" and "where to look" simultaneously under visual field constraints and examined whether learning was feasible under these conditions. As a result of training, a comparison of the number of blocks broken in the early stages of learning showed that the agent with visual field constraints successfully broke more blocks than the agent without constraints. Despite the imposed visual limitations, these results confirmed that it is possible for an agent to learn both in-game actions and gaze movement actions at a certain level.

In the second environment, a bullet-hell shooting game, we examined how introducing visual field constraints influenced in-game behavior. By connecting the left and right edges of the screen within the game, we created an environment where agents without visual constraints could move freely without considering the screen boundaries, whereas agents with visual constraints found it difficult to see both edges simultaneously and had to act more cautiously. To implement the visual constraints, we provided the agent with two types of input: a globally blurred image covering the entire screen and a cropped 3×3 image centered on the gaze area. In contrast, the agent without visual constraints received two different inputs: a full-screen image without blur and an image centered on the agent where the opposite edges of the screen are seamlessly connected. To compare behavioral differences between these agents, we counted the number of times they crossed the screen boundaries. After training each model for 50 million steps, we found that the agent with visual constraints crossed the screen boundaries only about half as often as the agent without constraints. Additionally, heat maps representing in-game coordinates revealed differences in the behaviors of the two agents. These results confirmed that visual field constraints influenced in-game actions.

Although challenges remain, such as slower convergence in learning due to the addition of gaze selection actions and the fact that the gaze behaviors do not necessarily resemble that of humans, this work successfully implemented learning in an environment with visual field constraints and confirmed the acquisition of behaviors that take these constraints into account.

概要

近年，人工知能（AI）技術は，画像認識や自然言語処理といった様々な分野で研究が進められ目覚ましい発展を遂げている．そのベンチマークとしてゲームが使用されることがあり，主にその強さを競って AI 技術が研究されてきた．囲碁や将棋といった完全情報のボードゲームで AI が人間のプロプレイヤに勝利しただけでなく，リアルタイムで仲間と協力して戦うようなデジタルゲームでも人間のプロプレイヤにチーム戦で勝利しており，ゲーム AI 研究の 1 つの目標であった人間よりも強いゲーム AI が多くのゲームで実現されている．しかし，強さを追求した AI のふるまいはしばしば人間らしさとはかけ離れることがあり，これを人間プレイヤの対戦相手としたり，教師とするには様々な課題が生じる．そのため，人間らしいゲーム AI を実現するための研究が進められている．

人間らしいゲーム AI を達成するためのアプローチとして，人間プレイヤに共通している動作や認識に関する制約を持つエージェントを用いて，強化学習を行う手法が存在する．これは，人間が持っているような認識精度や反応速度の限界を導入した環境で学習させることで，AI 特有の超反応や高精度の入力を必要としない，人間らしい挙動を獲得させることを目指している．本研究では，これまであまり着目されていなかった,「はっきりと見える範囲が限られている中で，その範囲を見たい場所に動かしながら行動する」という人間の認知・行動特性に着目した．

本研究は，視野による制約を前提とした行動を学習させることを研究目的とし，強化学習エージェントに動的な注視点を持たせるアプローチを提案した．このアプローチの特徴は，注視点からの距離に基づく視覚情報の制限と，その注視点の動的な制御である．強化学習エージェントは，情報を正確に認識できる範囲が限られ，その範囲も自ら制御しなければならない環境下で学習を行うことになる．2 つの異なるジャンルのゲームに対して視野の制約を導入した実験を行い，このアプローチを検証した．

一つ目の環境であるブロック崩しゲームでは，入力画像へのぼかしが強化学習エージェントに対して視覚的なハンデギャップになることを確認した後，視野の制約を導入した環境での強化学習を行った．具体的には画像をぼかす手段としてブラーを採用しブラーによるぼかし処理を施さない設定と，ボールが認識できる程度の弱いブラー，ボールが認識できないほどの強いブラーの 3 つを用意した．各設定 3 試行ずつ強化学習を行った結果，ブラーの強度に応じて性能が低下することが分かった．これにより，ブラーを用いたぼかし処理が，エージェントへの入力として視覚的なハンディキャップとして機能することが示された．

次に，同ゲームの画面を複数のエリアに分割し，エリアごとに強度の異なるブラーをかけた環境で強化学習を行った．ゲーム画面を 7 × 7 分割し，注視エリアを中心とする 3 × 3 エリアにブラーをかけず，その外側の 7 × 7 エリアに弱いブラーを，それよりも外側には強いブラーをかけることで，視野の制約を模倣した．注視エリアは，上下左右に動かすことができるようにし，この行動と環境内行動の組み合わせでエージェントの行動を定義した．視野の制約を持つエージェント

が「ゲーム内でどう行動するか」と「どこを見るべきか」を同時に学習しなければいけない環境を作成し，この状態で学習が可能かどうかを確認した．学習の結果，学習序盤で壊したブロックの数を比較すると，視野の制約を持つエージェントの方が，制約のないエージェントよりも多くのブロックを壊すことに成功していた．視野の制約があるにも関わらずこのような結果が得られたことから，エージェントに視野の制約を課したうえで，ゲーム内行動と視野移動行動の両方を一定のレベルで学習させることができることを確認した．

二つ目の環境である弾幕シューティングゲームでは，視野制約の導入がゲーム内でのふるまいに違いを生むことを確認した．画面上の左右をゲーム内で繋げることで，視野制約が無ければ画面の端を意識せずに移動できるのに対し，視野の制約を持つプレイヤにとっては画面の両端を同時に見ることが難しいため，慎重な行動を取る必要がある環境を作成した．画面全体に弱いブラーをかけた画像と，注視エリアを中心とする 3×3 エリアを切り取った画像の 2 つを入力とすることで，視野の制約を持ったエージェントを作成した．視野の制約を持たないエージェントは，ブラーがかかっていない画面全体の画像と，画面端でも反対側が含まれるような自機を中心とする画像の 2 つを入力とした．これらのエージェントがゲーム画面上の端を飛び越えた回数をカウントすることで，視野制約の有無による，ゲーム内のふるまいの違いを比較した．各モデルを 5000 万ステップ学習させて比較すると，視野の制約があるものは，視野の制約がないものの約半分しか端を飛び越えないという結果が得られた．また，ゲーム中の座標を表すヒートマップからも，両者のゲーム中のふるまいに違いが見られた．これらの結果から，視野の制約がゲーム内の行動に影響するということが確認できた．

視野選択行動の追加により学習の収束が遅く性能が低いことや，見ている場所が必ずしも人間と似ているとは限らないなどの課題は残るものの，本研究では視野の制約を導入した環境での学習を実現し，それを前提とする行動の獲得が確認できた．

目 次

図 目 次

表 目 次

第1章　はじめに

近年，人工知能（AI）技術は目覚ましい発展を遂げており，画像認識や自然言語処理，コンテンツの生成など，様々な分野で実用化が進められている．AI 技術の研究において，明確なルールと目標を持つゲームは，そのベンチマークとして広く利用されてきた．獲得した得点や AI 同士の勝敗といったわかりやすく比較できる結果を競う形で，AI 技術が進歩してきた．

これまでのゲーム AI 研究では，主に強さの追求に重点が置かれてきた．Deep Q-Network に代表される深層強化学習 [2] や AlphaGo でも使用されたモンテカルロ木探索 [3] など様々な手法が開発されたことにより，1 つの目標であった人間よりも強いゲーム AI が多くのゲームで実現されている．2015 年には DeepMind が開発した AlphaGo[3] が囲碁のプロ棋士に勝利を収め，2019 年には OpenAI の AI システム [4] が Dota 2 の世界トップレベルのプロチームに勝利している．また 2020 年には DeepMind による Agent57[5] が，Atari2600 に収録されている 57 種類のゲームにおいて，人間の平均的なスコア以上のスコアを収め，ゲーム AI の技術が汎用的に発展していることを示した．特に将棋や囲碁といったボードゲームでは，AI は既に人間のトップレベルの棋士を圧倒する強さを獲得しており，人間プレイヤーが AI を用いて学習することは一般的となっている．

しかし，こうした強さを追求した AI の振る舞いは，しばしば人間らしさとはかけ離れることがある．例えば，ボードゲームにおける人間よりも深く高精度の読みや，格闘ゲームにおけるフレーム単位の完璧な反応など，AI は人間の能力をはるかに超える性能を示す．このような人間を超越した振る舞いは，対戦相手として違和感を生じさせるだけでなく，学習の手本としても適していない．例えば将棋 AI の指す手は，その後何十手先までミスがないことを前提とすれば最適な選択かもしれないが，多くの人間プレイヤにとっては危険な選択となるかもしれず，真似するべきとは限らない．このように，AI の人間とは異なる挙動は実用面で様々な課題を生じさせるため，人間らしいゲーム AI を実現するための研究が進められている．

人間らしい AI の実現に向けた研究にはいくつかのアプローチが存在する．例えばルールベースで AI の挙動を制御し，人間らしい工夫を詰め込む手法や，人間プレイヤーのデータを利用した教師あり学習などは直感的な手法である．しかし，これらには膨大な工数が必要だったり，大量の学習データが必要だったりという課題が存在する．アプローチの 1 つに，人間プレイヤに共通する，動作や認識における基本的な性質に着目し，強化学習の際にそれらを考慮した環境で学習を行う手

法がある．このアプローチでは，人間が持っているような反応速度の限界や認識精度を導入した環境で学習させることで，AI 特有の超反応や高精度の入力を必要としない挙動を獲得させ，人間らしい AI の実現を目指す．また，このような研究はゲームという適用対象において人間らしいゲーム AI を作るだけでなく，AI 目線では非効率的な人間特有の行動がどういう原因で生じているのか，人間そのものを理解することにもつながる重要な研究であると考える．

本研究では生物学的制約の中でも特に，人間の視覚特性の一つである中心視野と周辺視野の違いに注目した．人間の視野には，網膜の細胞分布に起因する特徴があり，視線の中心から離れるほど視力が低下する．この特性により，人間は視野全体を均一に認識することができず，注意を向けた部分とそうでない部分で情報の取得精度に大きな差が生じる．そのため，視界に映るすべての情報を同時に正確に認識することは人間には困難であり，人間は意識的に，あるいは無意識のうちに重要な情報が得られそうな箇所に注意を向けている．得られる情報が不完全であることをふまえ，人間はそうした状況下でも大きな問題が生じないようなふるまいを獲得している．視野の制約を導入することで，AI 目線では非効率的かもしれない行動だが，人間にとっては自然な行動の実現を目指す．

本論文の構成は以下の通りである．第 2 章では関連研究として，ゲーム AI における人間らしさの研究や，人間の視覚特性について詳しく述べる．第 3 章では提案手法として，人間の視野特性を模倣する AI の概要について説明したのち，それによって期待する効果について，具体的な例を用いて説明する．第 4 章では比較的単純な環境であるブロック崩しゲームを用いた実験について説明する．第 5 章では，より複雑な環境である 2D シューティングゲームにおいて実施した実験について説明し考察する．最後に第 6 章では，本研究の成果をまとめるとともに，今後の課題と展望について述べる．

第2章　関連研究

本章では，関連研究として，まずは人間らしいゲーム AI に関する，本研究とは異なるアプローチについて，先行研究や実際の事例を述べる．次に，本研究の直接の関連研究として位置づけられる，生物学的制約に関する研究を述べる．最後に，人間の視覚特性やその再現についての研究を取り上げる．

2.1　人間プレイヤのログを直接活用する人間らしいゲーム AI

人間らしいゲーム AI を実現するためのアプローチは様々存在する．この節では主に，本研究とは異なるアプローチの先行研究や事例について述べる．

直感的なアプローチの一つに，人間プレイヤのログをそのまま教師あり学習の教師データとする方法が存在する．ゲームの状態とその時人間が選択した行動のセットを教師データとして学習することで，新たな状態が与えられた際に人間らしい行動を出力できるようにする．Ortega らによる研究 [6] や Maia[7] などがその代表例である．McIlroy-Young らによる Maia は，チェスにおいてオンラインサービス上での人間の棋譜を教師あり学習し，人間の着手を模倣するチェス AI である．棋力帯ごとに分けられた棋譜 1200 万局をそれぞれ学習させることで，各棋力帯に特化したモデルを作成した．盤面の探索を制限した強いチェス AI や学習途中の強化学習 AI と比較しても，探索を用いない Maia の方が人間の着手を高い精度で予測できることが示された．また，Maia は探索を用いないため，人間でも少し読むと悪いとわかるような手を指してしまうことがあるが，小川らは強化学習による強い AI が出力する確率分布と Maia の確率分布を組み合わせる Blend モデルを用いることでこれを緩和し，より高い精度で着手を予測することに成功している [8].

実際のゲーム開発における人間らしいゲーム AI の事例として，ストリートファイター 6 に実装されているまねもん機能を取り上げる．格闘ゲームのゲーム AI には，複数の役割が期待される [9]．操作方法の習得を支援したり，キャラクタ特有の動きを見せて魅力を紹介したりといったチュートリアルのような役割を担うゲーム AI は，必ずしも人間らしくある必要はない．一人用アクションゲームの強い対戦相手として，行動パターンを推測して対策を考える楽しさを提供する場合においても，対戦相手が人間らしい行動をする必要はない．一方で，駆け引きの理解を促す役割や，対人戦の予行演習にするためには，対戦相手となるゲーム AI がある

程度人間らしい行動をとることは重要である．各ランク帯のプレイデータから作成される「まねもん」機能は，Maia と同様に，ランクマッチ機能で収集した人間プレイヤ同士の対戦履歴を教師あり学習し，ランク帯ごとの AI を提供している．さらに，プレイヤ自身の履歴を追加で学習させることで，自分を模倣する AI と，同じキャラクタ同士で対戦することができる．

しかし，こうした教師あり学習による模倣には，元となる人間のプレイログが大量に必要となる点が課題になる．チェスやオンラインゲームのように学習データが豊富に用意できるとは限らず，新規のゲームやプレイヤの母数が少ないゲームでは適用が難しい．このような課題に対して，人間プレイヤのログを直接は利用せず，人間の特徴を明示的にゲーム AI に組み込む手法が存在する．次節では，こういった手法のうち，生物学的制約を学習時に実装することで人間らしい挙動の獲得を目指す研究について詳述する．

2.2 生物学的制約を用いた人間らしいゲーム AI

藤井らは，認知・行動の揺らぎや遅れといった人間プレイヤに共通して生じる現象を，生物学的制約として定義し，強化学習に導入した [10]．その結果スコアやタイムに最適化された行動ではなく，余裕を持って敵を跳び越すジャンプや，敵が多い場面では安全になるまで待つといった人間らしくみえる行動を獲得させることに成功している．この研究は制約を前提とした行動獲得という点で，本研究の直接の先行研究である．

藤井らの研究では画面全体に対して均一に揺らぎをかけていたが，平井らはシューティングの AI に人間の視覚特性を取り入れ，自機から離れた情報ほど強いゆらぎをかけることで人間らしいゲーム AI の作成を目指した [11]．均一なゆらぎではなく，距離に応じたゆらぎをかけることによって，視野の中心部と周辺部の情報精度の差を再現したことは重要だが，平井らの研究では注視点が自機に固定されている．人間プレイヤは必ずしも操作キャラクタだけを見ているわけではなく，また敵キャラクタや敵の撃つ弾のどれかだけを注視しているわけでもない．実際の人間プレイヤはゲーム画面の様々な場所に視線を移動させ，状況に応じて注視する箇所を変えているはずであり，ゲーム内の挙動もそれに影響されると考える．この気づきが，本研究の着眼点である．

2.3 人間の視野と視力に関する研究

本節では，我々の着眼点に関係する，人間の視野や視力に関する研究を紹介する．人間の視野は中心視野と周辺視野に分けることができ，それぞれ異なる特性を持つことが知られている．網膜の中心部分である中心窩とその近傍に映る範囲は中心視野，それ以外の網膜周辺部に映る範囲を周辺視野とされ，それぞれの領

域における網膜上の視細胞の分布密度が異なる．そのため，中心視野は比較的高い空間解像度を有し，詳細な情報を捉えることができるが，周辺視野は動きの検出は可能だが中心視野と比べると極端に視力は低い [1].

小松原は，人間の視界からの認知について，視野の限界や弁別能力を詳細にまとめている [1]．また，単に中心視野と周辺視野の二分だけでなく，視野内での視力の変化や，弁別能力の段階的な違いも指摘している．図 2.2 は視線からのずれに応じた視力の変化を示すグラフであり，図 2.1 は視野角度ごとの弁別能力を小松原が体系的にまとめたものである．

視力の低下をブラー（ぼかし）処理によって再現する研究も存在する．森は，公共空間のバリアフリーを点検するための室内疑似体験システムに対して，ガウシアンブラーによる画像処理を組み込むことで，視力低下レベルに応じたロービジョン（矯正視力でも 0.5 未満の状態）再現環境を開発した [12]．このシステムでは，ガウシアンブラーによるぼかしと，ランドルド環を用いて計測する視力を対応付け，各視力でどのように見えるのかを再現している．中心視の視力低下は水晶体の調節が上手く行えないことによって発生するため，カメラレンズの焦点のずれによるボケを再現するガウシアンブラーを用いることで，視覚能力レベルに対応する画像を作成することができる．周辺視野の視力低下は網膜上の視細胞の密度分布に起因する可能性が高いため，ガウシアンブラーによる再現は必ずしも十分とは言えないものの，本研究でもブラー処理による視力低下の再現手法を採用している．

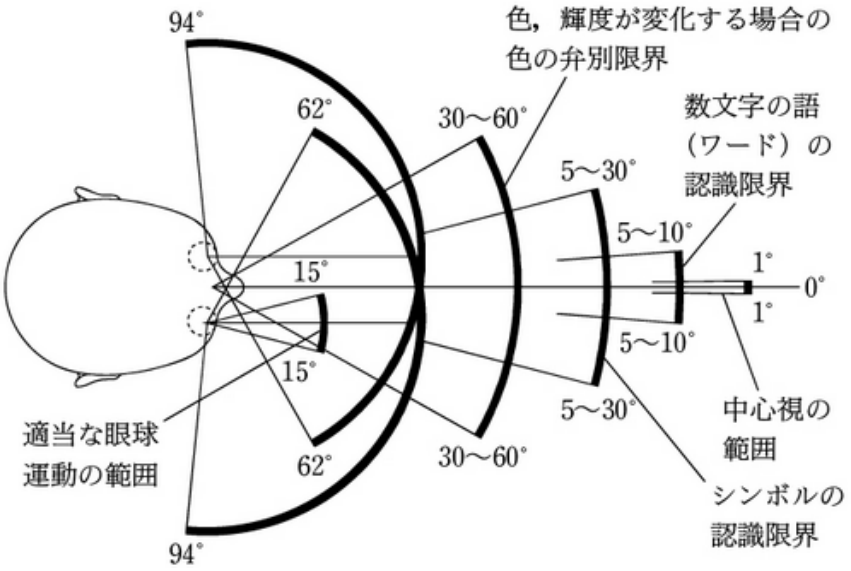


図 2.1: 視野と弁別能力　[1] より引用

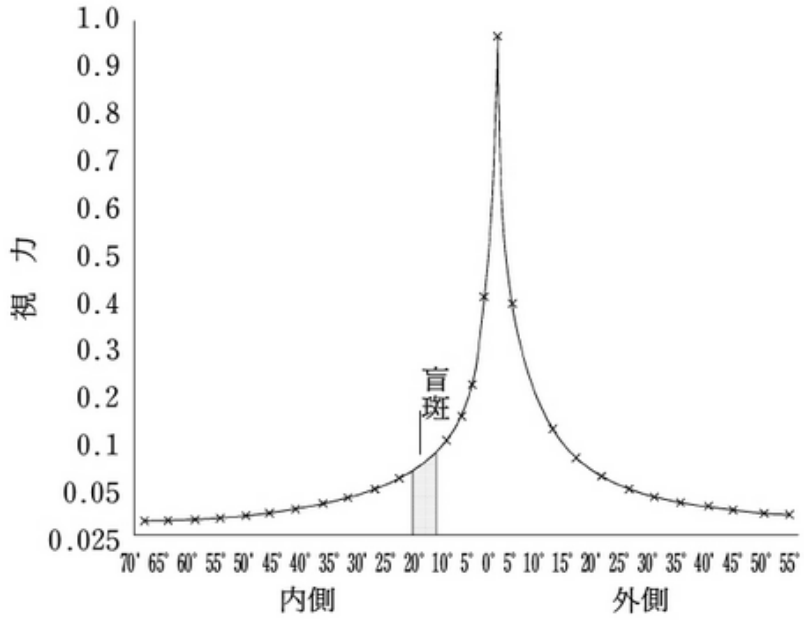


図 2.2: 視線からのズレと視力　真島英信，“生理学”，p241，文光堂（1978）より引用

第3章　アプローチと期待する結果

本章では，まず視野の制約を前提とする行動について具体的な例を用いて説明する．その後，これを実現するために，本研究が用いるアプローチと期待する結果について説明する．

3.1　視野による制約を前提とする行動

関連研究で述べたように，人間の視野は場所によって空間解像度が異なるという制約が存在する．この制約により，視野の中心部では高い解像度で情報を得られるが，周辺部では詳細な情報を得ることが難しい．このような視野の制約を前提として人間が行っている行動を再現することは本研究の一つの目的である．この節では，人間が視野の制約を前提として行っている行動について具体例と共に説明する．

視野の制約を前提とした行動には，次の 2 つの要素が含まれる．1 つ目は，不完全な情報を補うための能動的な行動である．人間は視覚から得られる情報をもとに行動するが，視野の制約によって見間違えや見落としが発生してしまう．それを補いより良い視覚情報を得るために，危険個所を注視したり，確認しやすいように行動したりすることがある．例えば，運転中にバックミラーやサイドミラーを定期的に確認する行動や，そのために徐行・停止する行動のように，視野の制約によって生じる不完全な情報に対して，能動的に情報を補おうとする行動が 1 つ目の特徴である．

2 つ目は，空間解像度の低下が重大な問題とならないような行動の選択である．視野の制約により不完全な情報しか得られない環境下でも，詳細な情報を必要としない行動を選択することができれば，視野の制約に起因する事故の発生確率を下げることができる．また，発生しうる事故が，より影響の少ないものになるような行動を選択することもある．例えば，運転手が左折する際に車を左側に寄せることで存在を見落としやすい左後方からすりぬけようとする自転車との接触リスクを低減させる行動が挙げられる．視野の制約による情報不足を補うのではなく，少ない情報でも安全な行動を選択することで対応するという点が，1 つ目の行動との違いである．

こういった行動の具体例として，2D 格闘ゲームにおける「置き」と呼ばれる行動について取り上げる．2D 格闘ゲームでは，相手の急速な接近や強力な攻撃には

多くの場合それぞれ適切な対処方法が存在する一方，適切な対処のため注視するべき場所がそれぞれ異なっていたり，またよく似た動作で紛らわしかったりと，すべてに対処することは難しい．そこで，相手の行動を認識する前に，相手が攻撃範囲に入ってきた際に当たるように技を出しておく,「置き」と呼ばれる行動が用いられる．相手が動いていない場合「置き」は空振り，その隙に反撃を受けるリスクが存在する一方で，いくつかのメリットも存在するため有効な選択肢として用いられている．相手が攻撃するモーションを素早く認識することが可能となれば，その攻撃を防ぐことができる可能性を上げることができるため，相手キャラクタを注視することは視野の制約の下では重要である．一方で，相手キャラクタだけを注視していると，相手のジャンプ攻撃への対応が遅れたり，残り時間や体力といった重要な情報を上手く得ることができなかったりする.「置き」はこういった問題に対応することができる．攻撃発生から技の出戻り（硬直）の間，キャラクタは動くことができない．そのため,「置き」の少し前に相手が行動していたり，硬直に相手の攻撃が当てられそうになっていても，どうすることもできない．また,「置き」が機能し，相手に攻撃が当たっていた場合についても，攻撃成功とともに発生するエフェクト（画面や音響の効果）によって相手に攻撃が当たったことを容易に確認することができる．つまり,「置き」を行っている間は，相手キャラクタの状態を常に注視する必要性が低下するため，その時間を利用して他の重要な情報を得るために注視点を移動させることが可能となる．また，相手キャラクタを注視するデメリットの一つに，相手のジャンプの方向を見誤ることで大きな隙を晒してしまう可能性が挙げられる．この課題に対しても,「置き」を実行しながら空中を注視することで，方向を見誤り対処を間違えるリスクを軽減することが可能となる．そもそも「置き」を見てから対処することが難しいことも相まって，しばしば「置き」は有効な選択肢として用いられる．このように，視野の制約を前提とした「置き」の戦略は，一定のリスクを伴うものの，注視点の効率的な活用と行動選択を組み合わせることで有効な戦略として機能する．これは本研究が目指す，視野制約を前提とした行動の代表的な例であり，人間は視野の制約の下で効果的に機能する戦略を見出している．

3.2　アプローチとその効果の予想

3.1 節で挙げたような行動を実現するために，本研究では強化学習エージェントに動的な視野を持たせるアプローチを提案する．このアプローチの特徴は，次の 2 点である．

1 つ目は，注視点からの距離に基づく視覚情報の制限である．画面内のある座標を注視点として定義し，その点からの距離に応じて取得できる情報の詳細さを変化させる．これにより，人間の視野に見られる中心部と周辺部での情報の解像度の違いを表現する．2 つ目は，注視点の動的な制御である．エージェントは，ゲーム内での行動に加えて注視点の移動も選択する必要がある．

このアプローチは，様々なゲーム環境に適用可能な枠組みである．エージェントは，情報を正確に認識できる範囲が限られ，その範囲も自ら制御しなければならない環境下で学習を行うことになる．通常の強化学習と比べて利用可能な情報が限られており，さらに視野を移動させるための行動も学ぶ必要があるため，学習の収束に時間を要したり，最終的な性能が低下することが予想される．しかし，3.1 節で述べたような人間の視野制約を前提とした行動の獲得が期待出来る．例えば，視野の制約により見落としが発生する可能性がある状況では，エージェントは重要な場所を注視したり，見落としのリスクを低減させる行動を選択したりすることが考えられる．

第4章　ブロック崩しへの視野制約導入

本研究では，2つの異なるジャンルのゲームに対して視野の制約を導入した実験を行った．この章では比較的単純なゲームであるブロック崩しでの実験について述べる．まず，4.1節では対象とする環境について，先行研究や本実験で使用した共通の設定について記載する．続く4.2節では画面全体にブラーをかける実験を行い，エージェントにとってブラーが視覚的な制約の一つとなることを確認する．最後に，視野の制約が存在する環境でも強化学習エージェントを学習させることができるか調べた実験について4.3節で述べる．

4.1　ブロック崩しと強化学習手法

本実験では，OpenAIの提供するgymnasiumに含まれている，Breakoutを題材とした[13]．Atari Breakoutは1976年Atari社によって開発されたブロック崩しゲームであり，プレイヤーはパドルを左右に動かしてボールを打ち返し，画面上部に配置された複数のブロックを破壊することを目的とする．ブロックを全て破壊するか，ボールを落としてライフを全て失うとゲームが終了する．

Breakoutはこのような簡素なゲームルールを持つため，様々な強化学習手法のベンチマークとして使用されてきた．Mnih et alの提案するDeep Q-Network (DQN) [2]ではニューラルネットワークを用いてゲーム画面を状態として直接使用し，Breakoutを含む複数のAtariゲームで人間レベルの性能を達成した．このDQNをベースとして，Rainbow[14]やAgent57[5]をはじめとする多くの手法が開発され，今ではBreakoutを含む複数のAtariゲームで人間平均的な点数よりも高い点数を達成している．

本研究では，強化学習エージェントの行動を環境内行動と視野移動行動の組み合わせで定義するため，通常よりも行動の種類が増加する．そのため，行動空間が大きい場合でも学習が可能である，Dueling Network[15]を用いた．Dueling Networkでは状態の価値と各行動の相対的な優位性を独立して推定することで，行動数が多い環境でも効率的に学習することができる．本実験ではWang et al. [15]と同一のハイパーパラメータを用いた．

4.2　ブラーによる性能低下の確認

2 章では，人間の視野は場所により視力が異なることと，視力が低下している状態をぼかしの画像処理によって再現した研究があることを述べた．強化学習でも同じように，ブラー等を用いて解像度を落とした画像をネットワークの入力に与えることができる．しかし，それが強化学習エージェントにとってハンデギャップになり，性能の低下を引き起こすのかについては自明ではない．

この節では，ブラーの強度が異なる複数の設定を用いて実験を行い，視覚的なハンデが学習性能にどのような影響を与えるかを検証する．

4.2.1　実験設定

本実験では，視覚的なぼかし処理が強化学習エージェントにとってハンデとして機能するかを検証する．エージェントの性能に対するぼかしの影響を明確にするため，ぼかしの強度が異なる 3 つの条件で比較実験を行った．実験条件として，ぼかし処理を施さない標準の状態，ボールが認識できる程度の弱いブラー，ボールが認識できないほどの強いブラーの 3 つを用意した．

入力画像には，[2] や [15] と同様にゲーム画面を 84×84 のグレースケールに変換したものを使用する．これを図 4.1 のようにぼかして実験を行った．図 4.1 左の画像がぼかしのかかっていない状態であり，中央がボールが認識できる程度のブラー（以下弱いブラー）を，右がボールが認識できないほどのブラー（以下強いブラー）をかけた画像である．これらのぼかしには OpenCV の Blur 関数を使用した．

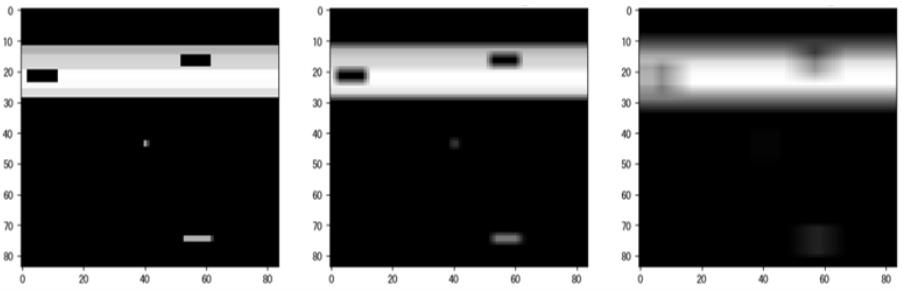


図 4.1: ブラーの強弱による入力画像の違い

4.2.2　結果と考察

図 4.2 は，各設定における性能（壊したブロックの数）の推移を示したものである．各設定 3 回の試行を行い，3 試行の平均を実線で，その 95%信頼区間を帯で表

した．強いブラーを書けたものは試行ごとのばらつきが激しいが，ブラーをかけ学習したものは標準のものに比べて性能が落ちていることが分かる．ボールが認識できない強いブラーの 1 つの試行では，弱いブラーの平均に迫る結果を出しているが，大まかに認識できる壊れたブロックを手掛かりとする挙動を学習しようとしているものと推測している．

この結果から，画面全体に弱いブラーや強いブラーをかけることが，エージェントに対して制約になることが確認できた．次節からそれぞれを中心に近い周辺視野に対応する加工，中心から離れた周辺視野に対応する加工として実験を行うこととする．

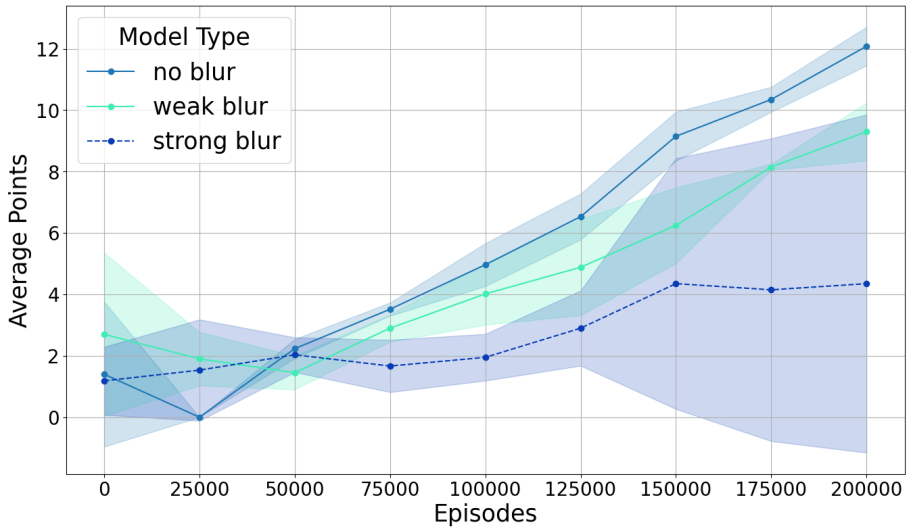


図 4.2: ブラーの強弱による性能差

4.3 視野制約と注視エリアの移動を導入した強化学習

ここからは，強化学習エージェントに視野の制約を実装し行った実験について記載する．前節の結果を踏まえて，強度の異なるブラーを用いて視野の制約を近似し，この環境下で視野の移動を含めた学習が可能かどうかを確認する．

4.3.1 実験設定

本研究の着眼点は人間の視野内の分解能の違いであるが，これを正確に再現し実装することは困難である．そこで，本実験では画面全体を 7 × 7 のエリアに分割したうえで，その中の 1 つである注視エリアからの距離に応じた，強度の異なるブラーをエリア単位で適応させることで，視野を模倣することにした．図 4.3 は，

具体的なブラーのかけ方を示した図である．注視エリアを中心として 3 × 3 エリアにはぼかしをかけず，その周囲 2 エリア外側の 7 × 7 エリアには弱いブラーが，それよりも外側は全て強いブラーをかけている．次の 3 種類の設定を用いて，視野制約の違いによる学習への影響を検証した．

1. 視野制約のない標準的な設定（以下，標準設定）
2. 視野制約が存在し，注視エリアがボールの存在するエリアに自動的に移動する設定（以下，ボール注視設定）
3. 視野制約が存在し，注視エリアを強化学習エージェントが上下左右に移動させることができる設定（以下，自由注視設定）

自由注視設定のエージェントは，3 章で述べた視野制約のアプローチに基づいて以下のような行動決定プロセスを実装した．まず，エージェントは自身が注視するエリアを番号として保持し，この注視エリアを中心に図 4.3 のようにぼかされたゲーム画面が入力として与えられる．次に，パドルの移動といったゲーム内で選択可能な行動と，注視点の移動という視野制約に基づく行動を組み合わせた行動空間から，1 つの行動を選択する．例えば,「パドルを左に動かしながら注視点を右に移動させる」といった複合的な行動が可能となる．エージェントの選択した行動によって次の注視エリアとゲームの状態が決定され，それに応じて加工された新しい画像が次の入力となる．このサイクルを繰り返すことで，エージェントが視野制約下での効果的な行動戦略を学習することを期待する．

自由注視設定では,「ゲーム内でどう行動するか」と「どこを見るべきか」を同時に学習する必要があるため，学習の収束に多くの時間がかかったり，性能が低下したりする可能性がある．そのため,「ボールを注視し続ける」設定を用意した．「どこを見るべきか」をこちらで指定することで，視野の制約を持ちつつも，ある程度妥当な注視移動方策を持っているエージェントを作成し，自由注視設定や標準設定と比較する．

これらの 3 つの設定をそれぞれ 3 試行ずつ実験を行った．視野の移動を含めた学習が可能かどうかを確認することが目的であるため，各設定間の差が見られ始めた 200,000 エピソード時点で学習を止めている．

4.3.2 結果と考察

図 4.4 に各モデルを比較したグラフを示す．橙色で示しているボール注視設定や，濃い緑の線で示している自由注視設定が，青い線で示している視野の制約が存在しない標準設定よりも，200,000 エピソード時点では多くのブロックを壊すことに成功している．最終的な性能については同等か，視野の制約がない標準設定が上回ることが予想されるが，少なくとも学習序盤時点では視野の制約が存在する設定の方が高い性能を示した．

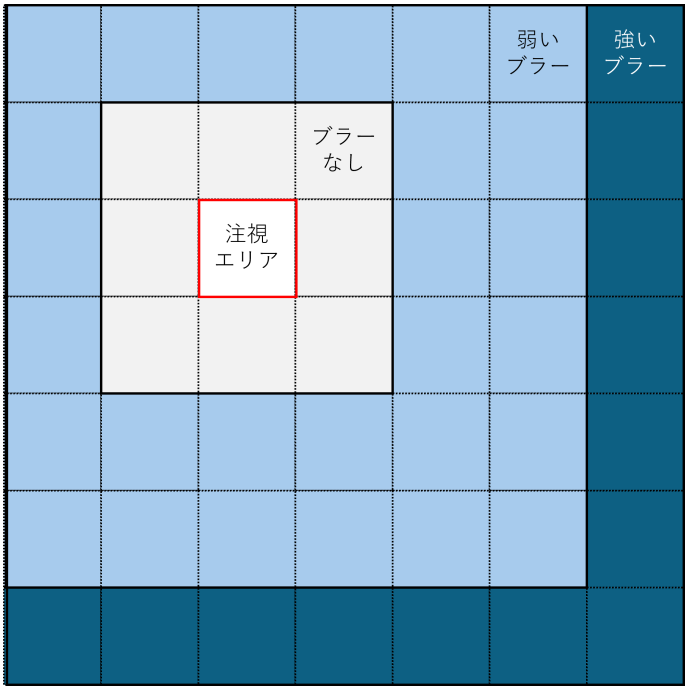


図 4.3: ブラーのかけ方

なお，エリアの分割を 3 × 3 とした実験も行った．注視エリアの 8 近傍に弱いブラーを，それよりも外側には強いブラーを適応した．結果を図 4.4 にて淡い緑で示しているが，この設定では上手く学習がされなかった．これは，注視エリアの移動に伴う入力画像の急激な変化が原因で発生していると考える．具体的には，3 × 3 分割では，注視エリアの移動前後で重複する非ブラー領域が存在しない．そのため，特にボールがエリアの境界線付近に存在する際に，その移動を認識することが難しく，エージェントの学習に悪影響を及ぼしたのではないかと考えている．

今回の実験では，エリアによって分割し，各エリアに異なるブラーをかけることでエージェントに視野の制約をかけた状態で，ゲーム内行動と視野移動行動の両方を学習させた．注視エリアがボールに固定される設定と，エージェントが注視エリアを自由に動かせる設定をそれぞれ学習させたところ，少なくとも序盤の性能については，どちらも視野の制約が存在しない設定よりも高くなることが確認できた．この結果から，エージェントに視野の制約を課したうえで，ゲーム内行動と視野移動行動の両方を学習させることが可能であると考える．

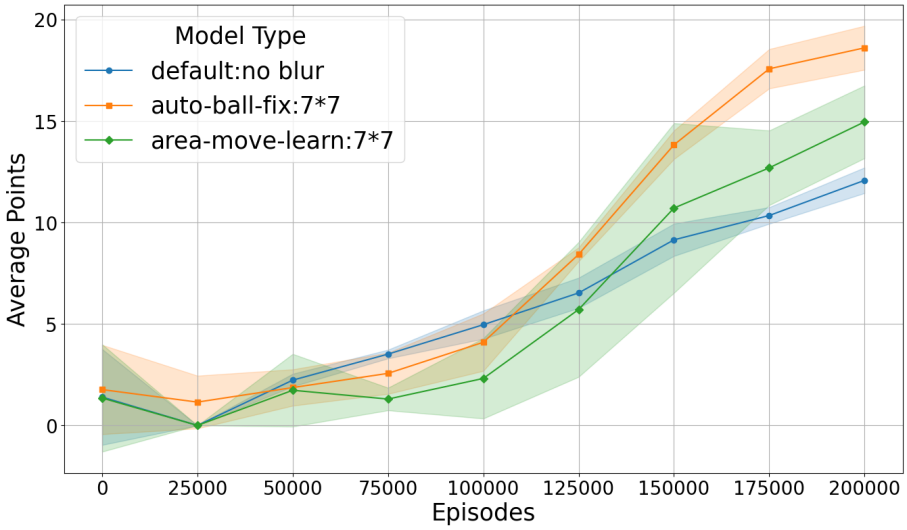


図 4.4: 画面を複数エリアに分割して行った実験

第5章　STG環境への視野制約導入

この章では，前章よりも複雑な環境である弾幕シューティングゲームに視野の制約を導入した実験について述べる．

5.1　対象とするシューティングゲーム環境と先行研究

本実験では藤本によって開発された弾幕シューティングゲーム環境を題材とした [16]．図 5.1 は実際のゲーム画面である．画面は白と黒の 2 色だけで構成され，このゲームに登場するオブジェクトは全て白で描画される．弾幕シューティングゲームとされているが，このゲームの目的は敵を撃ち倒すことではなく，敵の撃つ弾を避け続けることである．敵は画面上部に三角形で描画され，円形の弾を発射する．自機は正方形で描画され，その場にとどまるか上下左右に動かすことができる．自機を示す正方形の周囲 1 ドットのいずれかが白で描画されると被弾したみなされ，ゲームが終了する．

藤本はこの環境で強化学習エージェントの性能を上げる際に，入力に対する 2 つの工夫が有効であると示した．1 つ目は，全体を 84×84 に圧縮した画像だけでなく自機の周囲 84×84 ドットを切り取った画像も入力に含めることで，2 つ目は全体の画像を直近 4 フレーム分を重みをかけてまとめた 1 つの画像としていることである．図 5.2 が，実際に 4 フレーム分を新しいものほど白く描画されるようにまとめた画像であり，1 枚の画像から弾の方向や速度を判断することができるようになっている．この 2 つの工夫で，全体の画像で大まかな弾の方向を理解しつつ，自機を中心とする画像で細かな弾避けを狙っている．

解像度は粗いが画面全体が描画される画像と，解像度は高いが局所的な画像の 2 つの画像を入力とする構造は，視野の制約に近い構造であると本研究では考える．つまり，強化学習エージェントが自機を注視し続けている状態で学習していると捉えることができる．一方で，2 章でも述べたように人間プレイヤは状況に応じて様々な場所に注視点を移動させながら自機を動かしているはずである．そこで本研究では，藤本の作成したエージェントをベースとして利用し，局所的な画像の示す範囲を移動可能とすることで視野の制約を実装し，実験を行った．

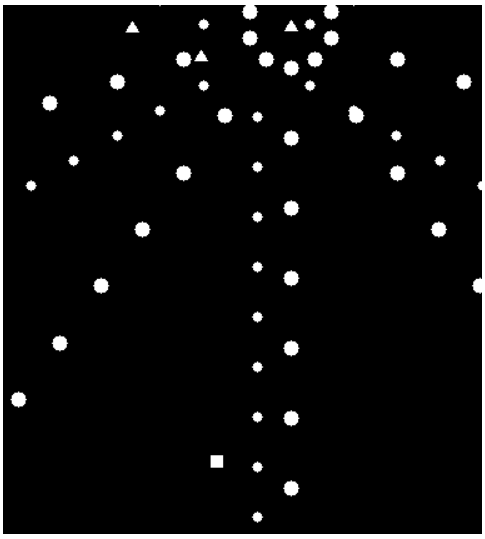


図 5.1: 藤本の開発した弾幕シューティングゲーム

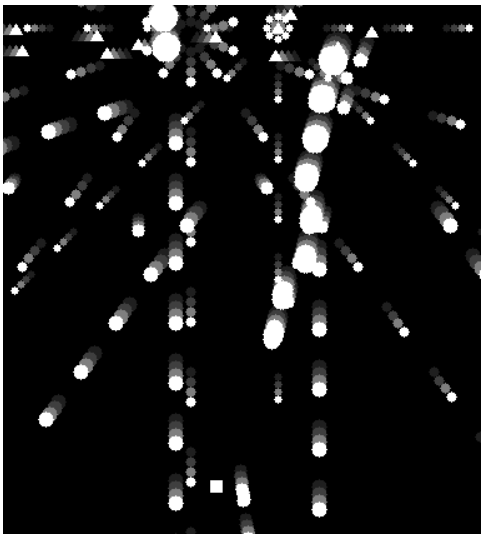


図 5.2: 4 フレームを重みをつけてまとめた画像

5.2 シューティング環境への視野選択行動の導入

本節では，視野制約の導入によってエージェントの行動に変化が見られることを確認するために，藤本の開発した環境を改変して行った実験について記載する．

5.2.1 実験環境の設計

藤本の研究の目的は詰み状況（弾や壁に囲まれ，この先どのように行動しても近いうちに被弾してしまう状況）を事前に回避するような AI を作成することであったため，詰み状況の発生しやすい，かなり難しい環境が開発された．この環境は視野制約がある場合には難易度が高くなりすぎること，人間らしい行動変化を生じさせる余地が小さいことから，いくつかの変更を行った．本項ではその詳細と意図について説明する．

変更後の環境において最も重要な点は，画面の右端と左端をゲーム内で繋げたことである．例えば自機が右端に位置するとき，さらに右に行動すると自機は左の端から出てくる．敵の弾も同様に，左右の端を超えると反対から弾が出現し，画面の下端を超えるまで消えることは無い．この変更により，自機が左右の端に到達し，これ以上横に移動して避けることができない，という状況が無くなる．本質的には画面上のどの x 座標に自機がいても同様の環境であり，弾との本質的な位置関係のみが行動選択に影響する．しかし，人間は画面の左端と右端を同時に見ることは難しいため，自機と弾の本質的な位置関係だけでなく，ゲーム画面上の x 座標も行動選択に影響するはずである．例えば，自機が中央から右端に向かって追い詰められているとき，そのまま右に進み続けて左の端に移動し（以後ワー

プとする）回避することができるが，両端を同時に認識することは難しく通常の移動よりも危険な行動となりやすい．また，左端付近から最短でこちらに向かってくる弾を認識しこれを回避することも難しい．そのため，この環境に少し慣れた人間プレイヤは端に追い詰められれば仕方なくワープを行うが，余裕があれば中央付近で弾をよけるようになると考える．

元の環境では画面外に出て消失していた弾が，左右を繋げる変更によって反対から出てくるため，元の環境よりも難易度が高くなる．視野の制約を導入することによる学習難易度の上昇も考慮し，敵の出現や発射する弾について，視認性と難易度のバランスを考慮した調整を行った．敵に関するパラメータを表 5.1 に示す．敵は自機を狙って弾を発射するものと，自機の座標に関係なく一定の方向に弾を

表 5.1: パラメータの変更前後の値

変数	意味	変更前の値	変更後の値
enemy_speed	敵の移動速度	5	5
enemy_shoot_timing_delay	敵が出てから静止し，弾を撃ち始めるまでの時間	21	21
bullet_speed_max	弾の最大速度	11	7
bullet_speed_min	弾の最小速度	3	1
bullet_radius_max	弾の最大半径	7	10
bullet_radius_min	弾の最小半径	3	5
bullet_num	発射される弾の最大数	11	15
bullet_delay	弾の発射間隔	8	25
n_way_max	最大発射方向数	12	6
probability	各フレームごとに敵が出現する確率	0.1	0.8
enemy_max	同時に出現できる最大敵数	無限	5

発射するものが存在するが，左右がループする環境に変更した関係でそのどちらも修正を行った．自機を狙って弾を撃つ敵は左右がループしていることを考慮し，道のりが最短になるように弾を発射するよう変更した．自機の座標に関係なく弾を発射するものは，1 〜 n_way_max の範囲内のランダムな方向に弾をばら撒く．各方向は 360 度を均等に分割して決定されるが，水平に発射され画面に残り続ける弾が出ないよう変更した．

5.2.2 実験設定

本実験では藤本の作成した Rainbow エージェントを参考に 3 つのモデルを用意した．この項では各モデルの詳細や，実験全体の設定について説明する．本実験

でも 4 章と同様に，画面全体を 7×7 のエリアに分割し，以下の 3 つの設定で実験を行った

名称	第 1 入力	第 2 入力	ブラー	注視エリア
視野無し設定	自機中央化した全体ラフ画像	自機中央化した自機周辺の画像	なし	-
自機注視設定	全体ラフ画像	注視エリア周辺の画像	全体に軽め	自機に追従＝自機周辺の画像
自由注視設定	全体ラフ画像	注視エリア周辺の画像	全体に軽め	隣接エリアへ移動可能

表 5.2: 各設定における入力画像と視野の設定

視野無し設定では，自機の画面上の位置にかかわらず行動させるため，自機の x 座標が画面の中心にくるよう再構成した画像をもとに入力を作成する．再構成した画像を 84×84 に圧縮した画像が 1 つ目の入力であり，2 つ目の入力は自機を中心として，他の設定と同じ範囲を切り取り，84×84 に圧縮したものとした．視野無し設定にも，場所により解像度の異なる入力しか得ることができないという制約はあるため，厳密には視野の制約を完全になくしたものではない．しかし，自機が左右の端付近に位置している際に，2 つ目の入力によって画面上では離れている反対側の様子を問題なく認識できることと，比較のベースラインとして扱うため，この設定を便宜上「視野なし設定」と呼ぶこととしている．

自機注視設定と自由注視設定では，通常のゲーム画面をもとに入力画像を作成する．全体の画像を単に圧縮するだけでなく，軽いブラーをかけた画像を入力とした．2 つめの入力には，注視エリアを中心とする 3×3 エリアを 84×84 にしたものを渡している．3×3 の範囲に画面外が含まれるとき，その範囲は敵や弾などと同じく白で塗りつぶしたものが渡される．

各モデルは 3 試行ずつ実験を行い，先行研究同様 5000 万ステップを終了条件とした．学習の各段階で 50 ゲームずつテストを行い，その時点での性能を記録している．元環境よりも本質的なゲームの難易度は低下した関係で，テスト中なかなか被弾せずいつまでもテストが終わらないという事態が発生した．そのため，テスト時には各ゲームで実時間に換算して 5 分が経過した時点で強制終了とした．

5.2.3 結果と考察

図 5.3 は，各設定における性能の推移を示したものである．視野無し設定は早い段階で性能を上げていることが確認できる．視野の制約が導入された 2 つの設定はこれと比べると性能の向上は緩やかではあるものの，学習が進んでいることが分かる．

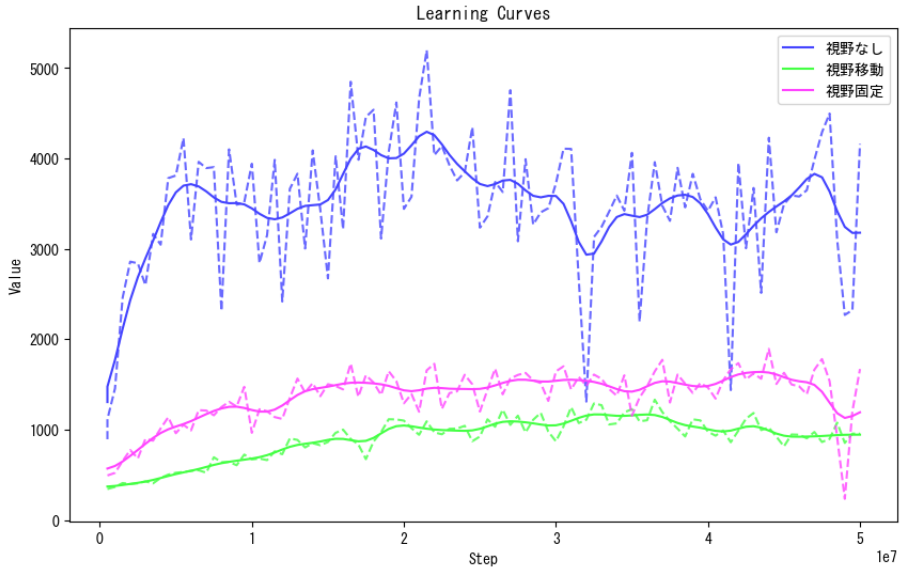
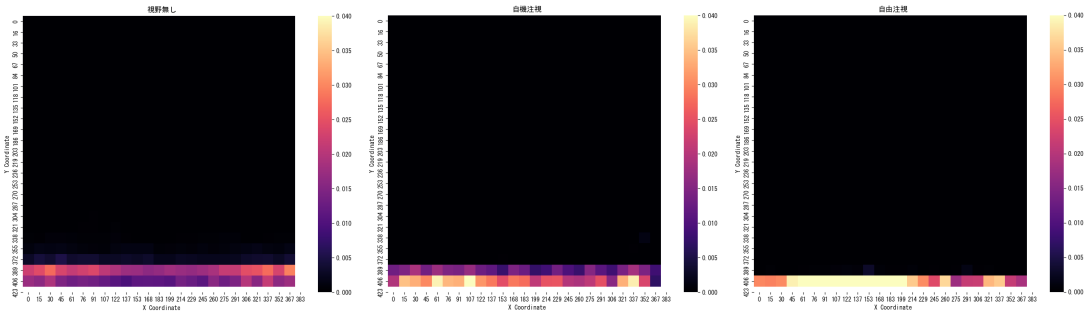


図 5.3: 学習曲線

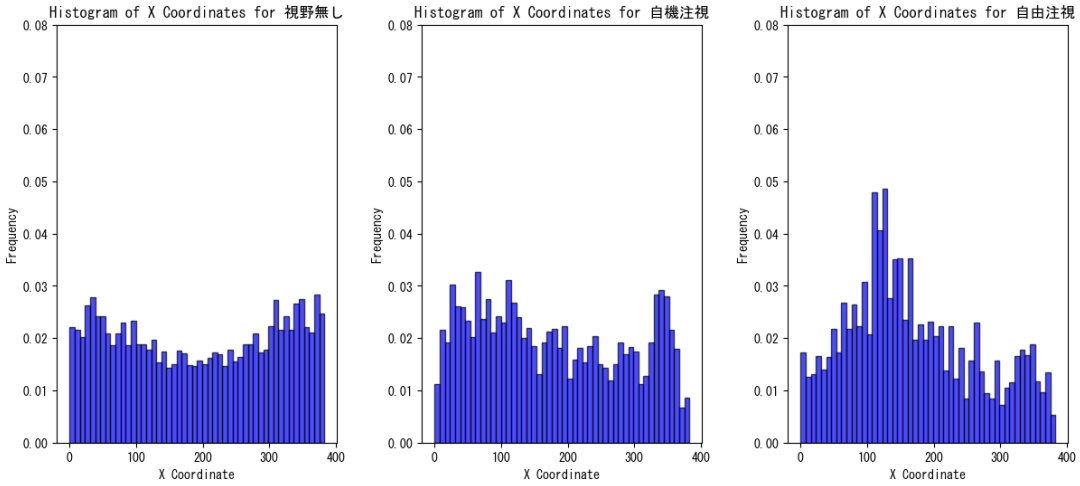
図 5.4 は各設定における自機の座標分布を示したヒートマップであり，このうち x 軸や y 軸だけに着目してヒストグラムにしたものが図 5.5 である．なお，ゲーム開始時の座標は常に画面中央下部であるため，各ゲームの最初の 3 秒間のデータは省いてヒートマップにしている．視野無し設定のヒートマップやヒストグラムからは，x 軸中央付近に滞在している時間よりも，画面の端の方に滞在している時間の方が長いことが確認できる．これは，敵の配置の関係で中央に弾が飛んでくる確率が少し高いためではないかと考察している．自機注視設定と自由注視設定のヒートマップから，視野の制約を導入したこれらの設定では基本的に画面下部に張り付くように行動していることが確認できる．



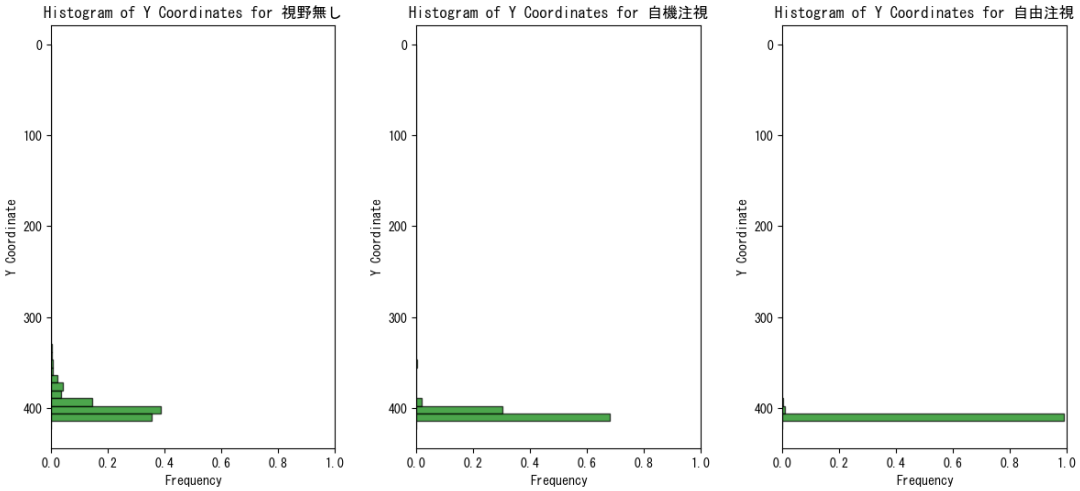
(a) 視野無し設定　(b) 自機注視設定　(c) 自由注視設定

図 5.4: 各設定の行動分布

図 5.6 は，各設定におけるワープ（画面左右の端を飛び越える移動）回数を比較した散布図であり．この統計量をまとめたものを表 5.6 に示す．視野無し設定が最



(a) 視野無し設定の x 座標　(b) 自機注視設定の x 座標　(c) 自由注視設定の x 座標



(d) 視野無し設定の y 座標　(e) 自機注視設定の y 座標　(f) 自由注視設定の y 座標

図 5.5: 3 種類の設定における x 座標・y 座標のヒストグラム

も頻繁にワープを行い，視野の制約がある設定と約 2 倍の差が出る結果となった．自機注視設定と自由注視設定を比べると大きな差はないが，少し自機注視設定の方が大きい．視野の制約が存在する設定ではワープ回数が少なくなっており，視野の制約を導入したことによりゲーム内の行動が変わることが確認できた．

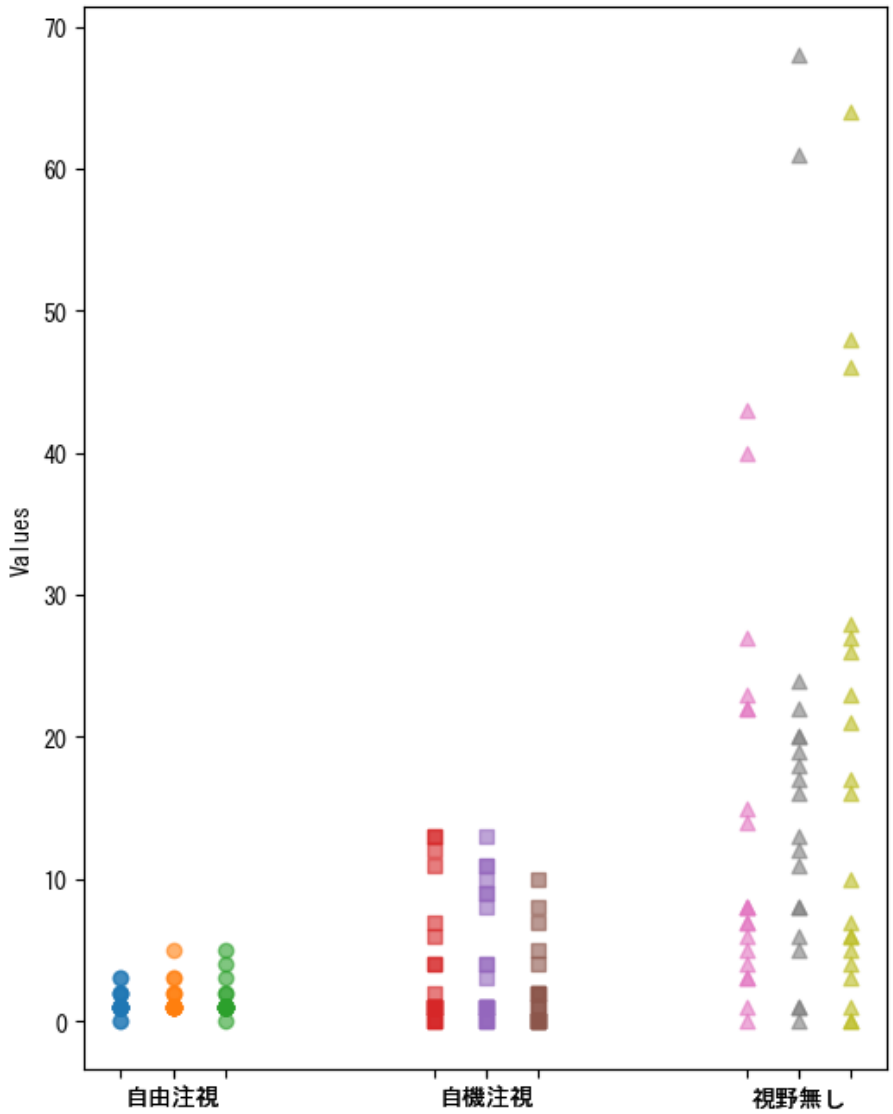


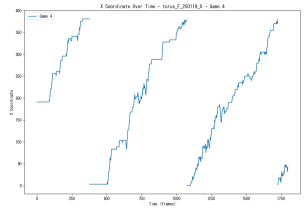
図 5.6: 各モデルが 10 秒間でワープした平均回数の散布図

具体的なゲーム内の動きについて，自機の x 座標の推移を示す図を図 5.8 にしめす．縦に x 座標を，横軸に時間を取り，あるゲームでの x 座標の序盤の推移を表したものである．視野なし設定では大きく 2 つのパターンが確認できた．1 つは図 5.8a のように，一定の方向に進み続けるパターンであり，もう一つは図 5.8b のように中央に広く空白がでる，端付近で立ち回るパターンである．自機注視設定に

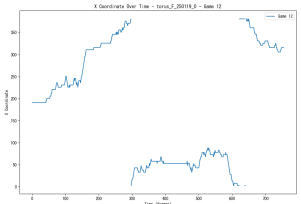
	Mean	Min	Max	Variance	SD
('Model', 'gaze_torus_F_250121_0')	0.559	0.0	0.852	0.049	0.221
('Model', 'gaze_torus_F_250123_1')	0.732	0.508	1.734	0.068	0.26
('Model', 'gaze_torus_F_250125_2')	0.532	0.0	0.742	0.024	0.155
('Model', 'gaze_fix_torus_F_250125_0')	0.572	0.0	1.907	0.297	0.545
('Model', 'gaze_fix_torus_F_250125_1')	0.793	0.0	4.212	0.87	0.933
('Model', 'gaze_fix_torus_F_250125_2')	0.368	0.0	0.873	0.107	0.327
('Model', 'torus_F_250119_0')	1.05	0.0	1.906	0.229	0.478
('Model', 'torus_F_250119_1')	0.924	0.0	1.665	0.146	0.383
('Model', 'torus_F_250126_2')	1.553	0.0	3.665	0.876	0.936
('Setting', '自由注視')	0.607	0.0	1.734	0.055	0.234
('Setting', '自機注視')	0.578	0.0	4.212	0.455	0.674
('Setting', '視野無し')	1.176	0.0	3.665	0.491	0.701

図 5.7: 各モデルが 10 秒でワープした平均回数

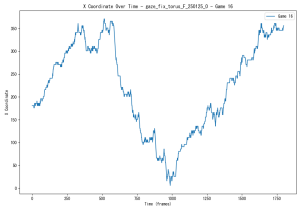
特有のものとして，図 5.8c のような端を避けるふるまいを確認することができた．視野の制約を持ったエージェントが，視野の制約がないエージェントには獲得されなかった行動を獲得したことを示している．



(a) ある方向に移動し続けるふるまい（視野無しに多い）



(b) 中央を避けるふるまい（視野なしの一部および，自機注視の一部）



(c) 端を避けるふるまい（自由注視や．自機注視に多い）

図 5.8: 各設定の行動分布

行動に変化が起こることは確認できた一方で，自由注視設定のエージェントの注視位置選択には更なる学習の余地が見られた．自由注視設定のエージェントが注視したエリアを示すヒートマップを図 5.9 に示す．

注視エリアの場所については更なる学習の余地が残る結果であるものの，視野の制約によってエージェントの獲得する行動が変化するという結果が得られた．

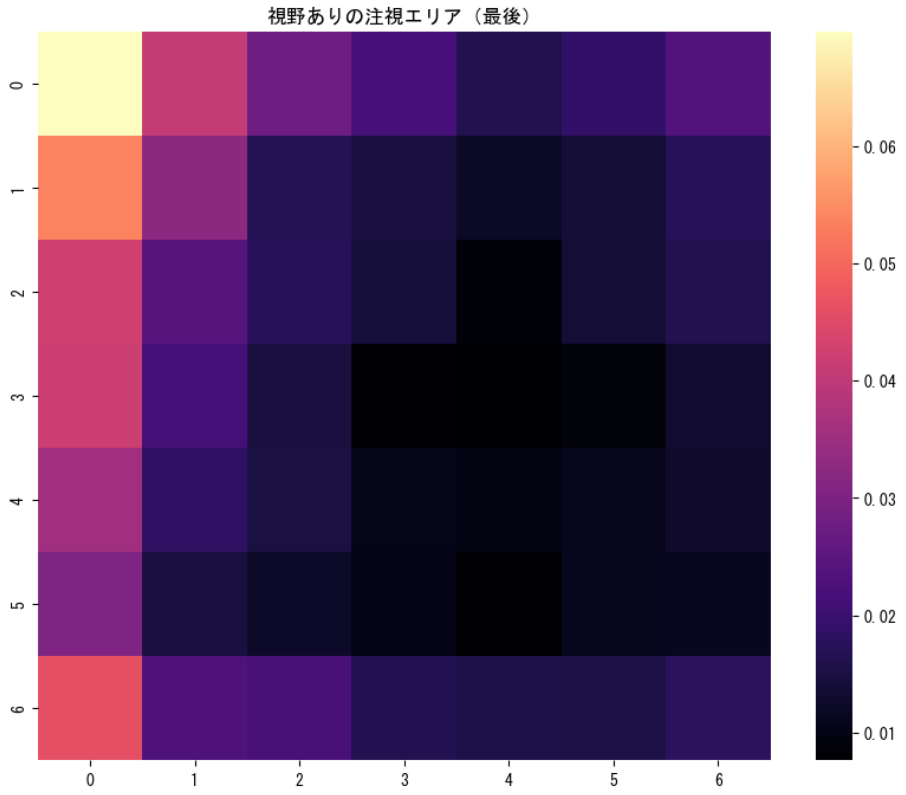


図 5.9: 自機注視が注視していたエリア

第6章　おわりに

近年の人工知能技術の進歩に伴い，人間よりも強いゲーム AI は多くのゲームで実現されてきた．一方で強さに特化したゲーム AI の挙動は人間らしさとはかけ離れることがあり，これを人間プレイヤの対戦相手や教師とするには様々な課題が生じる．そのため，人間らしいゲーム AI を実現するための様々なアプローチが提案されている．我々は，注視点があり，そこからの距離に応じて認識できる精度が変わるという人間の制約に着目した．この制約を前提とした行動を学習させるため，強化学習エージェントに動的な視野を実装するアプローチを提案した．

まず，ブロック崩しゲームを題材に，視野の制約を導入して強化学習を行うアプローチの実現可能性を検証した．入力画像に対してぼかしをかけた状態で強化学習を行った結果，エージェントにとってブラーによるぼかしの処理がたしかに認識の精度を低下させることを確認した．この結果を踏まえ，ゲーム画面を複数のエリアに分割し，注視エリアからの距離によって異なる強度のブラーをかけることで，視野の制約を模倣して，視野移動も含めた行動を強化学習エージェントに学習させた．実験の結果，注視エリア移動行動も含めた学習が可能であることを確認した．学習序盤で壊したブロックの数を比較すると，視野の制約を持つエージェントの方が，制約のないエージェントよりも多くのブロックを壊すことに成功していた．視野の制約があるにも関わらずこのような結果が得られ，学習の序盤において，視野の制約が性能向上に有利になる場合があることを確認した．

弾幕シューティングゲーム環境を用いた実験では，視野制約の導入がゲーム内でのふるまいに違いを出すか確認した．視野制約の導入による行動の変化をより明確に観察するため，左右の端が繋がった特殊な環境を作成した．この環境は，本質的には端が無く，画面上のどの x 座標にいても同様の条件であるため，自機と敵の弾の位置関係のみが自機の行動選択に影響する．一方で視野の制約を持つプレイヤにとっては両端を同時に認識することが難しいため，ゲーム画面上の x 座標も行動選択に影響するような環境である．この環境で強化学習を行うと，視野の制約があるエージェントは，視野の制約がないものよりも，端をワープして避ける回数が少なくなり，視野の制約によってゲーム内でのふるまいが変化することを確認することができた．エージェントが選択できる行動数が増えることで，学習の収束が遅くなったり，性能が低下してしまうという課題は残るものの，これらの実験を通じて，視野の制約を導入してもエージェントの学習は可能であり，視野の制約を前提とする行動が獲得されることが確認できた．

視野の制約を用いたアプローチによって，強化学習エージェントは視野の制約

を前提とした合理的な行動を獲得できると考えるが，これは人間の行動の再現にとどまらず，視野制約を前提とする未知の視野移動方策と行動を獲得できる可能性があると考える．このアプローチが，人間プレイヤの指導や，ゲーム開発（対戦ゲームのバランス調整やステージの自動生成）の支援といった分野へ貢献できることを期待する．

参考文献

[1] 小松原明哲. ヒューマンエラー. 丸善出版, 第 2 版, 2008.

[2] Volodymyr Mnih. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

[3] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, Vol. 529, No. 7587, pp. 484–489, 2016.

[4] Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung, Przemysław Dębiak, Christy Dennison, David Farhi, Quirin Fischer, Shariq Hashme, Chris Hesse, et al. Dota 2 with large scale deep reinforcement learning. *arXiv preprint arXiv:1912.06680*, 2019.

[5] Adrià Puigdomènech Badia, Bilal Piot, Steven Kapturowski, Pablo Sprechmann, Alex Vitvitskyi, Zhaohan Daniel Guo, and Charles Blundell. Agent57: Outperforming the atari human benchmark. In *International conference on machine learning*, pp. 507–517. PMLR, 2020.

[6] Juan Ortega, Noor Shaker, Julian Togelius, and Georgios N Yannakakis. Imitating human playing styles in super mario bros. *Entertainment Computing*, Vol. 4, No. 2, pp. 93–104, 2013.

[7] Reid McIlroy-Young, Siddhartha Sen, Jon Kleinberg, and Ashton Anderson. Aligning superhuman ai with human behavior: Chess as a model system. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1677–1687, 2020.

[8] Tatsuyoshi Ogawa, Chu-Hsuan Hsueh, and Kokolo Ikeda. More human-like gameplay by blending policies from supervised and reinforcement learning. *IEEE Transactions on Games*, 2024.

[9] 安原直宏. 『ストリートファイター 6』初心者から上級者まで対応した人間らしい有機的な行動を行う CPU（AI）のしくみ. CEDEC 2024. https://cedil.cesa.or.jp/cedil_sessions/view/2946.

[10] 藤井叙人, 佐藤祐一, 中嶌洋輔, 若間弘典, 風井浩志, 片寄晴弘ほか. 生物学的制約の導入による 「人間らしい」 振る舞いを伴うゲーム ai の自律的獲得. ゲームプログラミングワークショップ 2013 論文集, pp. 73–80, 2013.

[11] 平井弘一ほか. 弾幕の認識に人間の視覚特性を取り入れたシューティングゲーム AI の研究. ゲームプログラミングワークショップ 2016 論文集, Vol. 2016, pp. 158–161, 2016.

[12] 森一彦, 酒井英樹, 戒田真由美. 画像処理による視覚能力レベルに応じたロービジョン再現環境に関する研究. 日本建築学会計画系論文集, Vol. 76, No. 665, pp. 1213–1221, 2011.

[13] Mark Towers, Ariel Kwiatkowski, Jordan Terry, John U Balis, Gianluca De Cola, Tristan Deleu, Manuel Goulao, Andreas Kallinteris, Markus Krimmel, Arjun KG, et al. Gymnasium: A standard interface for reinforcement learning environments. *arXiv preprint arXiv:2407.17032*, 2024.

[14] Matteo Hessel, Joseph Modayil, Hado Van Hasselt, Tom Schaul, Georg Ostrovski, Will Dabney, Dan Horgan, Bilal Piot, Mohammad Azar, and David Silver. Rainbow: Combining improvements in deep reinforcement learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32, 2018.

[15] 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.

[16] 藤本修嗣ほか. 強化学習を用いた弾幕シューティングゲームを攻略するエージェントの作成. ゲームプログラミングワークショップ 2024 論文集, Vol. 2024, pp. 51–57, 2024.