

強化学習とSim2Realを用いたAI制御技術の製造装置への適用

仁田 脇 貴 浩¹・白 鳥 遼¹・葛 西 洋 平¹・田 畑 好 浩¹・黒 澤 公 紀¹

Application of Reinforcement Learning and Sim2Real to AI Control Technology for Manufacturing Equipment

T. Nitawaki, R. Shiratori, Y. Tabata, Y. Kasai, and K. Kurosawa

近年AI技術は目覚ましい進化を遂げ一般社会でも活用されはじめている。それに伴い製造業におけるAI技術の活用に対する期待も高まっている。当社では2015年からディープラーニングを用いたAI技術の適用に向けた活動を進めてきており、検査工程への導入などに成功してきた。本稿では装置制御の分野にAI技術を導入しAI技術の適用範囲を拡大した事例について説明する。

In recent years, AI technology has made remarkable advancements and has been increasingly utilized in various sectors of society. As a result, expectations for the application of AI technology in the manufacturing industry have also grown. Our company has been actively pursuing the implementation of AI technology using deep learning since 2015, and we have achieved successful integration in inspection processes, among other areas. This paper aims to describe a case study where AI technology was introduced in the field of device control, expanding the scope of its application.

1. ま え が き

近年のAI技術の進化は目覚ましく、一般社会にも急速に浸透が進んでいる。製造業においても例外ではなく、製造工程の自動化など、事業活動の中でAI技術を活用しようとする動きが活発となっている。

当社は、2015年からディープラーニングを用いたAI技術の工程適用に関する活動を開始し、2017年度には、外観検査AIの工程導入を実現した。この外観検査AIの導入により、不良品の後工程への流出や、検査にかかる工数を低減させる効果が得られている。

AI技術の工程適用を成功させるために、当社ではAIシステム開発ロードマップ（図1）を作成し、方針を明確にした上で段階的に取り組んできた。

図1において横軸にはAIが扱う情報、縦軸はその情報を扱う上でのキー技術が並んでいる。また情報とキー技術により実現可能な利用領域を棒グラフ内に記載している。AIが扱う情報を、“画像”、“数値・記号・音”、“動作制御”、“言語概念”の大きく4分類しており画像は“静止画単体”、“動画単体”、“複数”に、また数値・記号・音は“グラフ化”、“時系列”にブレイクダウンしている。これらは、技術的難易度の低い方から順に記載している。当社は、このAIシステム開発ロードマップに従いAIの製造導入の難易度が低い方から順に、適用を進めてきている。

本稿では、世界的にも実導入の事例が少ない製造装置の自律制御に関するAI適用事例を紹介する。以下、第2章では、キー技術である強化学習とSim2Realについて説明する。第3章と第4章では、取り組みの内容とその結果について説明し、最後に5章で本稿をまとめる。

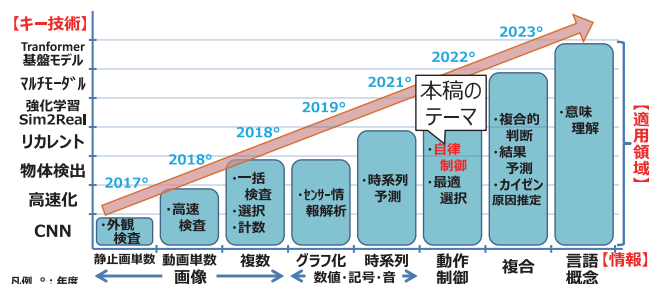


図1 AIシステム開発ロードマップ

Fig 1. AI system development roadmap.

略語・専門用語リスト

略語・専門用語	正式表記	説 明
AI	Artificial Intelligence	人工知能
CNN	Convolutional Neural Network	画像や動画認識に用いられるディープラーニングの一種
強化学習	Reinforcement Learning	現在の状況から取るべき行動を決定する問題を扱う機械学習手法
エージェント	Agent	強化学習において学習や行動を行う主体
Sim2Real	Simulated to Real	シミュレーションを用いて現実世界の課題を解く機械学習手法の一種

表 1 教師あり学習, 教師なし学習, 強化学習

Table 1. Supervised Learning, Unsupervised Learning, Reinforcement Learning

学習種類	①教師あり学習	②教師なし学習	③強化学習
概要	正解タグ(教師情報)がついたデータから学習する	正解タグを用いずに, データから学習する	行動に対する報酬が最大化するように学習する
メリット	期待通りの結果が得やすい	タグ付け不要, 取組やすい	試行錯誤で想定以上の結果
デメリット	タグ付け稼働が大	分析結果の良し悪しが人の分析能力に依存する	報酬設定が難, 学習時間大
主な用途	外観検査	データ分析	ゲーム

2. 用いた技術の説明

一般にAI技術は教師あり学習, 教師なし学習, 強化学習に分類される(表1)。本稿ではこれらのAI技術のうち装置制御に向く手法であると考えられる強化学習を用いたAI技術について説明する。

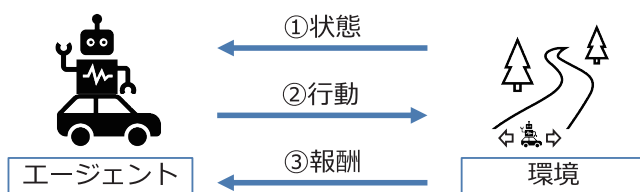


図2 強化学習のイメージ

Fig 2. Image of Reinforcement Learning.

2. 1 強化学習

強化学習は, エージェントが環境との相互作用を通じて適切な行動方針を獲得する学習手法である。

図2にイメージを示す。①環境は, エージェントに状態を伝える。②エージェントは, 状態に応じて行動を行い, 状態を変化させる。③エージェントは, 変化させた結果に

応じて報酬を受け取る。エージェントは, この流れの中で試行錯誤を行い報酬が最大になるよう最適な行動方針を獲得していく。

強化学習は, 行動に対して適切に報酬を設定できれば, 事前知識なしで学習を進めていくことができる。このため強化学習は行動や報酬が設定可能なゲーム分野での適用が進んでいる。装置制御においても行動と報酬が設定可能であるケースにおいて強化学習に向く手法であると考えられるが, 装置制御において, 実導入には課題も多く製造装置への適用事例は, まだ世界的にも珍しい。

2. 2 学習に関する課題

製造装置への強化学習の適用には, 大きく二つの課題があると考えられる。一つ目の課題は, 学習フェーズにおいて実際の製造装置を長時間使用する必要があり製造を制限する必要があること, 二つ目の課題は学習過程での試行錯誤により生じる意図しない動作により製造装置の故障や消耗のリスクが伴うことである。一般的に, ユーザは製造の制限や, 装置の故障や消耗は避けたいため, 製造装置での学習が難しい事も多い。これらが製造装置への強化学習が進まない要因となっていると考えられる。

この課題を解決するため, 本事例では強化学習の課題を

解決する方法として製造装置の使用することなく学習ができるSim2Realも併せて適用した。

2. 3 Sim2Real

Sim2Realはシミュレータ上で学習した強化学習AIを実環境へ適用する技術である。この技術を用いる事で、学習フェーズにおいて実際の製造装置を長時間使用する必要があり製造を制限する必要があるという課題と、学習過程での試行錯誤により生じる意図しない動作により製造装置の故障や消耗のリスクが伴うという課題を解決することができる。

一方で材料の形状や特性、センサやカメラへの不純物の付着など様々な要因で変化する実環境の状態やその捉え方をシミュレータで完全に再現することは難しく、一般的に実環境とシミュレータとのギャップが生じる。このギャップが大きい場合、そのまま製造装置への強化学習AIを適用すると意図しない動きになることがある。本事例でも実環境とシミュレータとのギャップ問題が発生しており、今回の取り組みを通じて解決してきた。これについて3.5節で説明する。

3. 製造装置への強化学習の適用

本章では第2章で説明した強化学習とSim2Realの適用例として、光源から放出される光のピントを調整するレンズの位置合わせに強化学習制御を適用した事例について説明する。

3. 1 概要

図3にレンズ調心の概要を示す。光源から出射するビームは拡がりをもっている。ビームを効率良く用いるために拡がりをなくした平行なビームに整形するため光源とレンズの位置と角度を調整する工程がある。この工程ではカメラで撮影したビーム画像から取得した各特徴量が製品としての規格値を満たすよう、主にY軸、 θ Y軸、Z軸、 θ Z軸の4軸を調心する自動調心するプログラム（従来自動調心）が導入されている。特徴量で最も重要なのはビーム幅であり、調心が最適な状態に近づくとなるとビーム幅は細くなっていく。

従来自動調心では1軸ずつあらかじめ定めた順番通りに調心を行っている。この方法だと本来調心する必要がない軸も毎回調心するため調心時間がかかる。また、レンズの微細な形状の違いや装置やその他の影響により調心中に意図せず軸がずれることがあるが、この方法だとその軸の調心の順番が終わっていれば、軸がずれたままでも調心完了

してしまい規格値を満たさない状態で調心完了してしまう。なお、規格値を満たさない場合は人が手動調心を実施し規格値に入るように再度調心を行っている。人が手動調心に対応する間は本来行うべき作業ができず負担になっている。

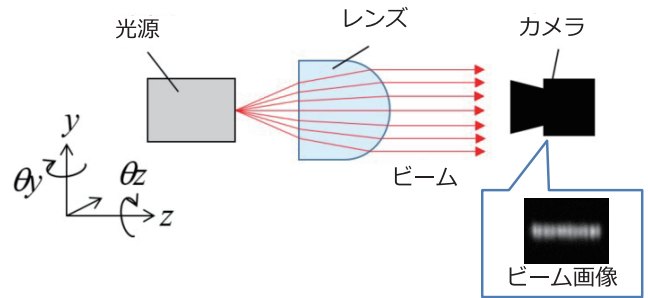


図3 レンズ調心の概要

Fig 3. Summary of Lens Alignment.

3. 2 シミュレータ

図4に従来のレンズ調心フローと、本稿で取り組んだ手法に基づく強化学習調心システムのシミュレータ構成を示す。従来のレンズ調心をシミュレートし、Y軸、 θ Y軸、Z軸、 θ Z軸の各軸に対する調心量に応じたビーム画像を生成するシミュレータを作成した。図5に実際のビーム画像とシミュレータによる生成画像の比較を示す。

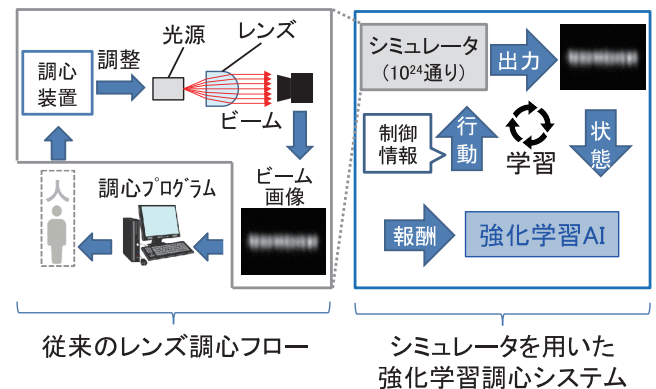


図4 従来レンズ調心フローとシミュレータの構成

Fig 4. Traditional Lens Alignment Workflow and Simulator Configuration.

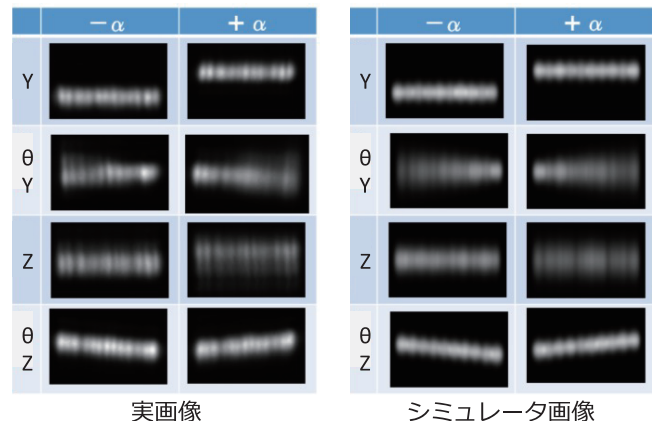


図5 実画像とシミュレータ画像比較

Fig 5. Comparison between Real Images and Simulator Images.

3.3 画像入力方式の強化学習AI

3.2節で説明したシミュレータ画像を用いて強化学習AIの開発を進めてきた。状態としての入力は図6のような単一画像である。①現在の状態を示す画像と、②現在の状態と1つ前の状態との差分画像、また画像入力を前提とした強化学習AIを活用するため、③画像化した制御情報を加え、1枚の画像に必要な情報をすべて含めるよう工夫を行った。制御情報の画像化に際しては、制御情報の次元数に合わせて画像を複数の領域に分け、各制御情報に応じた輝度を与える事で画像データとして表現した。以上のように過去の状態、現在の状態、制御情報をそれぞれグレースケール画像として表現し、RGB画像へ合成することで単一画像に複数の状態、および制御情報を盛り込んだ。報酬は、レンズ調心において最も重要な特徴量であるビーム幅とした。つまりビームが細ければ細いほど、より大きい正の報酬を与えた。さらに、調心回数に比例して負の報酬を与えることで、ビームが細くかつ調心回数が少なくなるような行動が高く評価されるように設計した。出力は次の行動で調心する軸とその軸の調心量とした。強化学習AIで調心する軸はZ軸、 θ Y軸、 θ Z軸とし、Y軸については他の軸との依存関係もなく従来自動調心と同じアルゴリズムで容易に調心可能であることから強化学習AIの対象外とした。

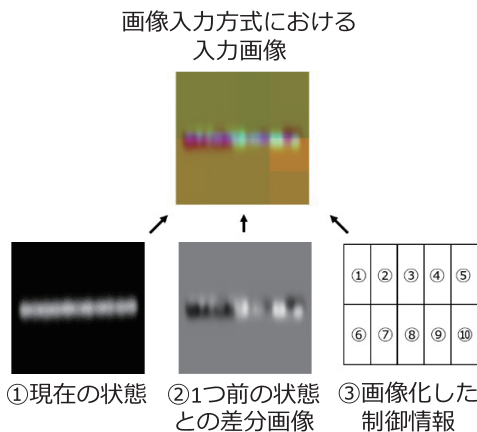


図6 画像入力方式における強化学習調心システムへの入力画像

Fig 6. Comparison between Real Images and Simulator Images.

上記の方法によりシミュレータ上で動作する強化学習AIの開発に成功することができた。この強化学習AIはビームの状態に応じて軸を適切な量調整するため従来自動調心より効率的に調心を実施することが期待される。しかし、強化学習AIを製造導入するにあたり2.3節で述べた実環境とシミュレータとのギャップに関する問題が発生した。

3.4 実環境とシミュレータとのギャップ問題

実際の製造装置に適用するにあたり、カメラへの微小な不純物の付着による干渉縞やレンズの軸がずれている場合に発生するサイドローブが発生し、実環境とシミュレータとの見え方のギャップが大きくなってきた（図7）。当初考えていた画像入力方式の強化学習AIを利用するとノイズなしと比較し特にサイドローブにおいて調心時間がかかることがわかった。（表2）。

この問題に対応するためには、干渉縞やサイドローブを反映したシミュレータを再作成して学習する必要がある。しかし、この干渉縞やサイドローブの見え方は製造装置やその日の状況などの様々な要因によっても変わることがわかっておりシミュレータへ正確に反映するのは困難である。

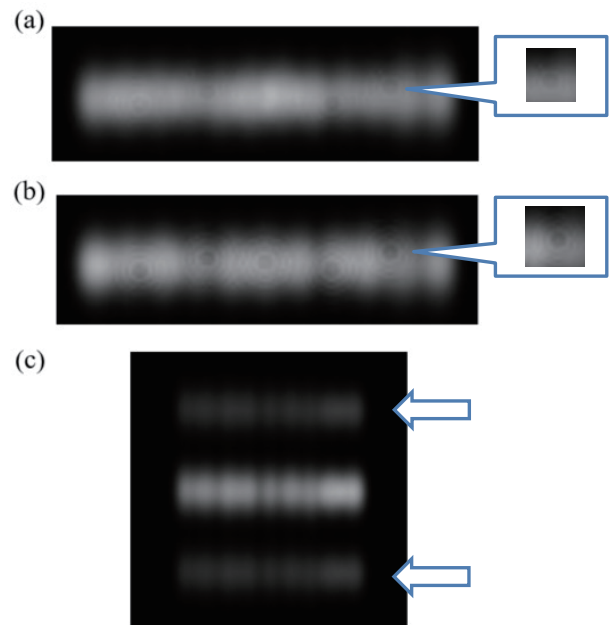


図7 画像例。

(a)干渉縞（弱）. (b)干渉縞（強）. (c)サイドローブ.

Fig 7. (a) Weak Interference Fringes.

(b) Strong Interference Fringes.

(c) Side Lobes.

表2 画像の見え方による影響

Table 2. Effect of image appearance.

入力画像種類	画像入力方式	
	調心回数	ビーム幅
ノイズなし	27.4回	13.2px
干渉縞（弱）	29.8回	13.2px
干渉縞（強）	29.6回	13.3px
サイドローブ	73.1回	13.4px

3. 5 特徴量入力方式での解決

実環境とシミュレータとのギャップ問題を解消するため画像から特徴量を抜き出して学習に用いる方法を考案した(図8)。画像からビーム幅やビームの傾きなどの状態を数値として強化学習AIに入力することで干渉縞やサイドローブなどのノイズの影響を抑えることができる。本方式は、製造装置やその日の状況などの様々な要因によって見え方が変わる実画像とシミュレータとの違いを吸収する効果が見込める。なお、報酬や行動は画像入力方式の強化学習AIと同じとした。

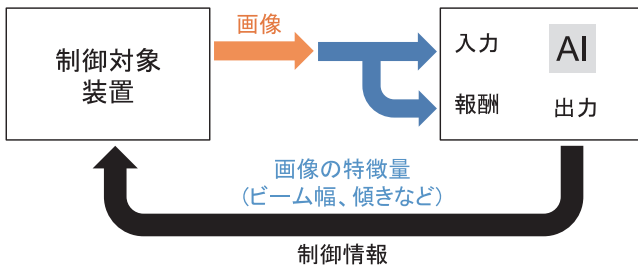


図8 強化学習の学習モデル

Fig 8. Learning Models in Reinforcement Learning.

表3に同じ条件で評価を実施した比較結果を記載する。両方式における調心回数、ビーム幅の平均値を示している。ノイズ付加なしの場合や干渉縞パターンの場合には、両方式に大きな差は見られないが、サイドローブの場合に特徴量入力方式では大きく改善した。

表3 画像の見え方による影響

Table 3. Effect of image appearance.

入力画像種類	画像入力方式		特徴量入力方式	
	調心回数	ビーム幅	調心回数	ビーム幅
ノイズなし	27.4	13.2px	28.8回	13.3px
干渉縞 (弱)	29.8	13.2px	29.5回	13.4px
干渉縞 (強)	29.6	13.3px	33.3回	13.5px
サイドローブ	73.1	13.4px	29.7回	13.1px

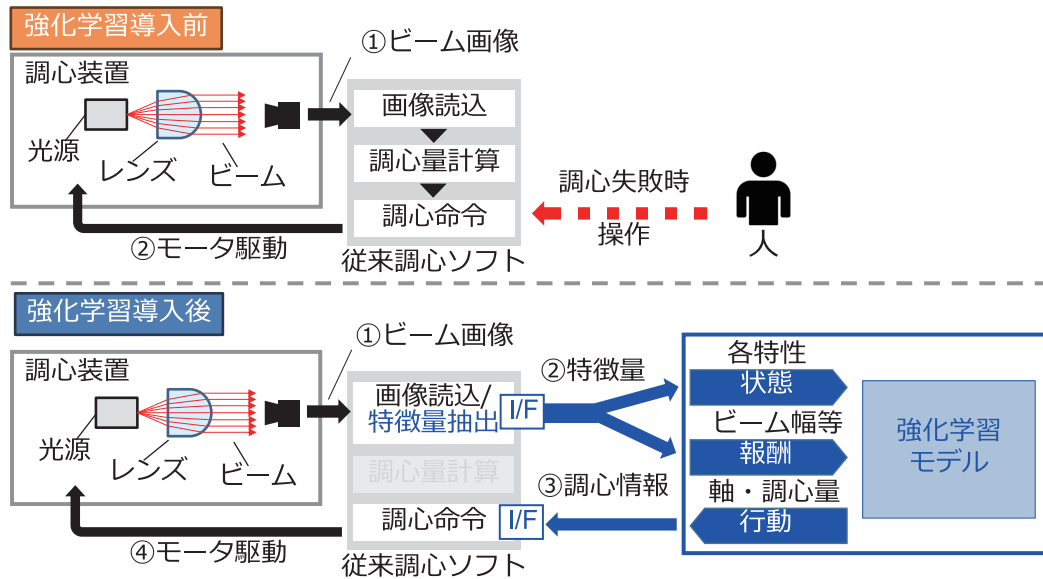


図9 強化学習導入前後のフロー比較

Fig 9. Comparison of Workflow Before and After Introducing Reinforcement Learning.

3. 6 実環境への強化学習適用

本項では3.5節で学習した特徴量入力方式の強化学習AIを実際の制御装置へ適用する取り組みについて説明する。図9に強化学習導入前後のフロー比較を示す。図中のI/Fはインターフェースを表している。3.1節で述べた通り、レンズ調心においては従来自動調心を実施する従来自動調心ソフトが存在する。強化学習AI導入にあたり、従来自

動調心ソフトにおいて画像読み部と調心命令部の2箇所のインターフェースを改造し、撮影画像の取り出しと、調心情報の受け取り機能を追加した。なお、シミュレータと同様、Y軸の調心は他の軸との依存関係がなく従来自動調心と同様のアルゴリズムで調心が可能であることから対象外とし、Z軸、 θ Y軸、 θ Z軸の3軸を強化学習AIの対象としている。

4. 評価

(1) 手動調心と強化学習調心の比較

実際の製造装置において強化学習調心の優位性を確認するため、手動調心、強化学習調心について、調心時間および調心結果の比較評価を行った。比較評価では、ベストな調心状態からZ軸を $\pm 2.5 \mu\text{m}$ 、 θY 軸を $\pm 0.25^\circ$ 、 θZ 軸を $\pm 0.5^\circ$ の範囲でランダムにずらした位置を初期状態とし調心を完了するまでの時間と完了後の調心結果について手動調心と強化学習調心でそれぞれ評価する。この評価では、異なる10種類の初期状態を用意し手動調心と強化学習調心で条件が同じになるように設定した。

図10に各軸の初期状態でのずれ量と調心時間の関係を示している。各グラフの横軸にある各軸のずれ量は、初期状態での各軸のずれ量を示しており、0の位置がベストな調心状態である。また、縦軸は調心が完了するまでの調心時間

間を示しており0に近い方が、調心時間は短い。プロットは初期状態での各軸のずれ量とその初期状態からの調心時間を示しており、手動調心は青丸、強化学習調心は緑丸で示している。この図から手動調心と比較し強化学習調心の方が短時間で調心完了していることがわかる。

図11に各軸の初期状態でのずれ量と調心完了時のビーム幅の関係を示す。このグラフも同様に、各グラフの横軸は各軸の初期状態での各軸のずれ量を示している。縦軸はその初期状態から調心完了した時点でのビーム幅をプロットしており、0に近い方がビーム幅は細いビームになっている。この図から強化学習調心は手動調心と比較して調心結果は同程度以上に良いことがわかる。2つの結果から、手動調心と比較し強化学習調心は、調心時間が短く同等程度以上の調心結果で調心できていることを確認できた。

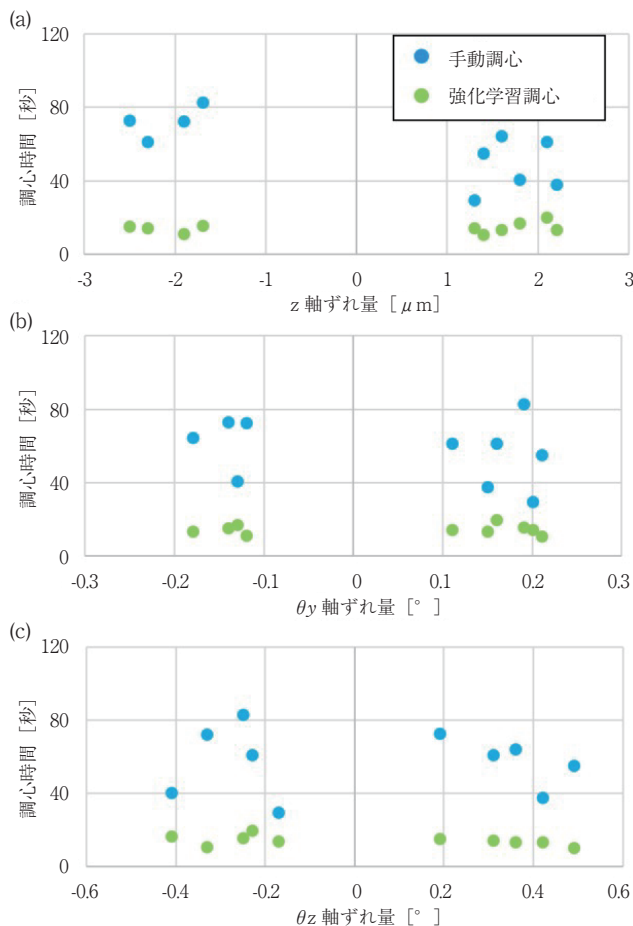


図10 各軸の初期状態でのずれ量と調心時間。

(a) z軸, (b) θy 軸, (c) θz 軸と調心時間。

Fig 10. Amount of Deviation and Alignment Time for Each Axis: (a) z-axis, (b) θy -axis, (c) θz -axis, and Alignment Time.

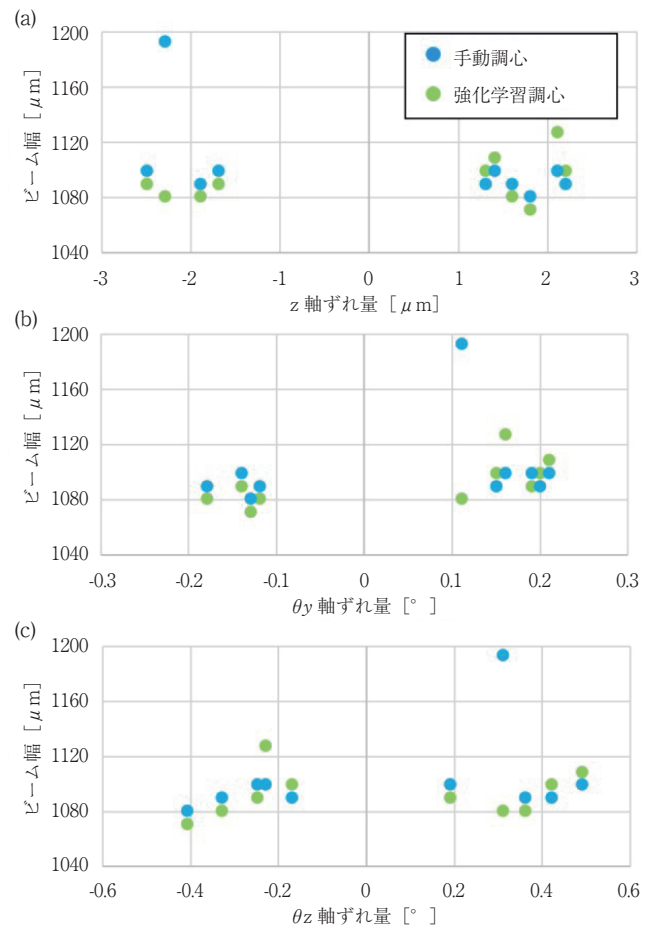


図11 各軸の初期状態でのずれ量と調心完了後のビーム幅。

(a) z軸, (b) θy 軸, (c) θz 軸とビーム幅。

Fig 11. Amount of Deviation and Beam Width after Alignment for Each Axis: (a) z-axis, (b) θy -axis, (c) θz -axis, and Beam Width.

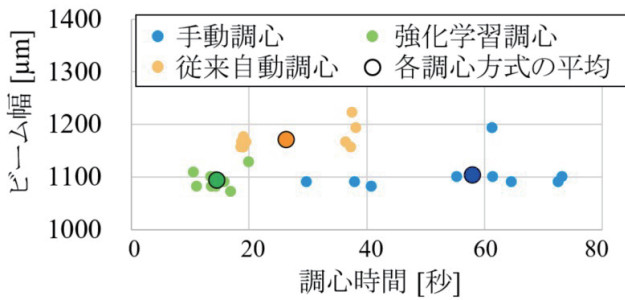


図12 実機における調心実験の調心時間とビーム幅の関係
Fig 12. Relationship between Alignment Time and Beam Width in Lens Alignment Experiments on the Actual System.

(2) 手動調心，従来自動調心，強化学習調心の比較
続いて，手動調心，従来自動調心，強化学習調心の調心時間と調心結果の比較を実施した．図12に調心時間とビーム幅を2次的にプロットした結果を示す．調心時間は短い方が良く，またビーム幅は細い方が良い調心結果と言えるので，この図において左下に近い点ほど短時間で高性能であることを表している．各調心方法の濃色マーカーは平均値を示しており，表4に結果を示す．調心時間について手動調心は57.9秒，従来自動調心は26.3秒，強化学習調心は14.5秒であった．ビーム幅について手動調心は1103.6 μm ，従来自動調心は1171.2 μm ，強化学習調心は1093.2 μm であった．強化学習調心は手動調心，従来自動調心と比較し，短時間かつ高性能になった．以上の事から，手動調心，従来自動調心と比較して，強化学習AIの優位性を確認することができた．

この取り組みを通し 2.2節で示した製造装置での学習に関する課題および3.4節で示した実環境とシミュレータとのギャップ問題を解決し，製造装置への強化学習AIの導入を達成した．また，3.1節で示した従来自動調心の課題を解消し，調心時間の短縮および調心結果の向上などといった効果が期待されている．

表4 実機における調心実験結果の平均

Table 4. Average Results of Lens Alignment Experiments on the Actual System.

	調心時間 [秒]	ビーム幅 [mm]
手動調心	57.9	1103.6
従来自動調心	26.3	1171.2
強化学習調心	14.5	1093.2

5. むすび

製造装置の自律制御に関するAI適用事例として，強化学習とSim2Realをレンズ調心に適用した事例を紹介した．製造装置へのAI技術の適用は世界でも珍しく当社でも初である．手動調心，従来自動調心と強化学習調心の比較を行い，強化学習調心が短時間で高性能であることを確認した．現在当社ではこの技術の製造装置への導入を完了し運用保守を実施している．今後も，更に多くのAI技術を適用し，事業上の効果を得ていきたいと考えている．

参考文献

- 1) 黒澤公紀, 仁田脇貴浩, チャン ドウック ソン, “AI化推進方法と，実現に向けた技術開発”，フジクラ技報, No.134, pp. 41 – 52, July 2021.
- 2) 白鳥遼, 仁田脇貴浩, チャン ドウック ソン, 葛西洋平, 黒澤公紀, “強化学習・Sim2Realを用いた製造業へのAI制御適用”，SSII 2022.
- 3) Tatsuo Akiyama, “Practical Use Case of Reinforcement Learning and Sim-to-Real Transfer for Manufacturing”，NVIDIA GTC 2023.
- 4) D. Silver, et al., “Deterministic policy gradient algorithms,” Proceedings of the 31st International Conf. on Machine Learning, vol. 32, pp. 387 - 395, June 2014.
- 5) V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg and D. Hassabis, “Human-level control through deep reinforcement learning,” Nature, vol. 518, pp. 529 - 533, Feb. 2015.
- 6) 高見 豪, “プラント制御AIの実現”，横河技報, Vol.63, No.1, pp. 33 – 36, 2020.
- 7) S. Daftary, J. A. Bagnell, and M. Hebert, “Learning transferable policies for monocular reactive mav control,” in International Symposium on Experimental Robotics, 2016.
- 8) J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, “Domain randomization for transferring deep neural networks from simulation to the real world,” in IEEE/RSJ International Conference on Intelligent Robots and Systems, 2017.
- 9) B. Balaji, S. Mallya, S. Genc, S. Gupta, L. Dirac, V. Khare, G. Roy, T. Sun, Y. Tao, B. Townsend, E. Calleja,

- S. Muralidhara and D. Karuppasamy, "DeepRacer: Autonomous Racing Platform for Experimentation with Sim2Real Reinforcement Learning," in IEEE International Conference on Robotics and Automation, 2020.
- 10) T. Mingxing and Le. Quoc, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Proceedings of the 36th International Conference on Machine Learning, vol. 97, pp. 6105 – 6114, 2019.
- 11) Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot and N. Freitas, "Dueling Network Architectures for Deep Reinforcement Learning," Proceedings of the 33rd International Conference on Machine Learning, vol. 48, pp. 1995 – 2003, 2016.
- 12) C. Xiao, P. Lu and Q. He, "Flying Through a Narrow Gap Using End-to-End Deep Reinforcement Learning Augmented With Curriculum Learning and Sim2Real," IEEE Transactions on Neural Networks and Learning Systems, 2021.
- 13) L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao and J. Han, "On the Variance of the Adaptive Learning Rate and Beyond," ICLR 2020.