Journal of Computer Technology and Applied Mathematics
Journal Home: www.jctam.online
Vol. 1, No. 3, 2024 | ISSN 3007-4126 (Print) | ISSN 3007-4134 (Online)

A Contrastive Deep Learning Approach to Cryptocurrency Portfolio with US Treasuries

LI, Zichao [1*] **WANG, Bingyang** [2] **CHEN, Ying** [3]

[1] Canoakbit Alliance Inc, Canada
[2] Emory University, USA
[3] MyMap AI, USA

***** *LI, Zichao is the corresponding author, E-mail: zichaoli@canoakbit.com***

Abstract: To enhance the portfolio performance of cryptocurrency and US treasuries, we introduce a pioneer approach that incorporates contrastive learning to traditional deep learning techniques. The contrastive learning process builds a contrary learning input whenever a market moves, or trader action is performed to enhance the learning supervision. We leverage contrastive learning with both Deep-Q Learning and Proximal Policy Optimization. The result shows promising performance in both downward and upward market environments.

Keywords: Cryptocurrency, Supervised Learning, Digital Asset, Bitcoin, Ethereum, Litecoin, Machine Learning, Crypto, Contrastive Learning.

DOI: https://doi.org/10.5281/zenodo.13357988 **ARK:** https://n2t.net/ark:/40704/JCTAM.v1n3a01

1 INTRODUCTION

This study is conducted in collaboration with a machine learning vendor for a cryptocurrency investment firm. The client has specified that the portfolio should be exclusively used to trade the following five cryptocurrencies: Bitcoin, Ethereum, Ripple (XRP), Litecoin, and Chainlink, along with their derivatives. The derivatives include futures, options on futures, and perpetual swaps. The buying and selling activities are financed through the repo of the firm's holdings in U.S. Treasuries, which are considered risk-free assets. Regular portfolio rebalancing is also required.

Past research efforts are mainly focused on arbitrary or profit taking strategies of cryptocurrency. There are numerous efforts investigating risk-free market instruments like US treasuries. However, there is no effort to discuss a balanced portfolio of risk assets of cryptocurrencies and risk-free assets of US treasuries. In this study, we would like to present a pioneer effort of implementing contrastive deep learning strategies in a balance portfolio of US treasuries and cryptocurrencies.

We will run solution approaches in different market conditions. Portfolio performance will be evaluated in terms of cumulative return, Sharpe ratio and Sortino ratio.

2 LITERATURE REVIEW

2.1 Cryptocurrency Portfolio Management

In the current financial market, portfolio optimization is one of the objectives to maximize returns of a portfolio [1]. A common strategy is called strategic asset allocation (SAA) that attempts to balance risks and returns with different weightages for target asset allocation.

Other than cryptocurrency physical coin product, we often rely on cryptocurrency derivatives to hedge the risk of crypto trading, the most used ones include future, options and swaps. Perpetual swap provides high leverage, and the trading is based on over-the-counter trading [2]. A cryptocurrency perpetual swap is a type of derivative contract that allows traders to speculate on the price movements of cryptocurrencies without actually owning the underlying asset. Unlike traditional futures contracts, perpetual swaps do not have an expiry date, meaning traders can hold their positions indefinitely as long as they maintain the necessary margin. To keep the swap price close to the spot price, perpetual swaps use a funding rate mechanism. This periodic payment between buyers and sellers is based on the difference between the perpetual contract price and the spot price. Perpetual swaps often offer high leverage, allowing traders to control a large position with a relatively small amount of capital. Traders can trade on margin, which means they can borrow funds to increase their position size.

The BitMEX XBTUSD perpetual swap is a highly traded derivative product that enables traders to speculate on the price of Bitcoin (XBT) relative to the US Dollar (USD) without the need to own underlying asset. For traders taking a long position, the underlying asset is Bitcoin. These traders anticipate that the price of Bitcoin will rise against the US Dollar. When this happens, the value of their long position increases, resulting in a profit. However, if the price of Bitcoin falls, the trader experiences a loss. Conversely, traders holding a short position also have Bitcoin as their underlying asset, but they expect the price of Bitcoin to decline against the US Dollar. When the price drops, the value of the short position rises, and the trader profits. If the price of Bitcoin increases, the trader incurs a loss.

2.2 MACHINE LEARNING FOR CRYPTOCURRENCY

[3] summarize all past literature on the application of machine learning in cryptocurrency research. [4] discuss the use of Random Forest and LSTM for cryptocurrency price predictions.

[5] investigates the time-series data modelling which extends from [6] [7]. Deep Reinforcement Learning is an approach that continuously learns from interactions to optimize the result. [8] proposes a deep Q-learning portfolio-management framework. Four cryptocurrencies are studies in the portfolio and the authors achieve the expected result. It has a local agent for assets and a global agent for global rewards. [9] designs a crypto portfolio management system from a deep neural network. [10] [11] combines LSTM models and reinforcement learning algorithms for cryptocurrency portfolio optimization. The algorithm is tested with a portfolio mixed with traditional commodity assets like SPY, wheat, WTI and gold. It does not include risk-free asset like US Treasuries. [12] proposes a Convolutional Neural Network (CNN) and Fully-Connect Neural Network (FCNN) approach for cryptocurrency portfolio management, which are by far the most popular approaches. However, none of those literatures above works on crypto derivatives like crypto futures, swaps and options. None of them investigates portfolios that need balance crypto instruments with risk-free-assets.

Most of those proposed approaches are not working well with cryptocurrency derivatives products. [13] [14] [15] propose attention LLM enlightens us to pursue positive and negative attention pairs like those used in contrastive learning. Contrastive learning is an instance-wise discriminative approach that aims at making similar instances closer and dissimilar instances far from each other in representation space [10]. The actual representation of the contrastive learning elements can be in generative AI form [16] or another standard readable format.

2.3 CONTRASTIVE LEARNING

Contrastive learning aims to learn representations by comparing similar and dissimilar pairs of data. In the context of portfolio management, this could involve learning to distinguish between different market conditions, asset correlations, or other relevant features. Contrastive learning is necessary because the same economic indicator can suggest different market conditions depending on the context. For instance, before the Federal Reserve decides to lower interest rates, a non-farm payroll report that falls short of expectations might boost the cryptocurrency market (leading to an increase in crypto values). However, after the Fed has already cut interest rates, a similar disappointing payroll report might discourage investment due to fears of an impending economic crisis.

2.4 REPO RATE AND RISK-FREE-ASSET

In our studied case, portfolio managers use US treasuries in the portfolio as collateral to fund the buy and sell of cryptocurrencies and their derivatives. This process is illustrated in [17]. Repo rates are usually reported daily based on actual transactions in the repo market. They can be published by financial data providers or central banks. Some organizations or financial institutions might calculate a benchmark overnight repo rate by averaging rates across multiple transactions or participants. Repo rates can vary based on market conditions, the quality of the collateral, and the term of the agreement. Rates can be obtained from financial market data providers or through direct quotes from financial institutions. The quality of the U.S. Treasury note as collateral can affect the repo rate you receive. Higher quality (e.g., shorter maturity, higher credit rating) can lead to better repo rates. The repo market is typically very liquid, but rates can fluctuate based on market supply and demand. Ensure you use up-to-date market rates for accurate cost calculations.

3 METHODOLOGY

Before we come to the actual deep learning model that we are going to apply, we first introduce some operational basics in the first two subsection.

3.1 REPO RATE AND FUNDING OF CRYPTO TRADING

Using U.S. Treasuries to fund cryptocurrency cash, derivatives like perpetual swaps involves leveraging the liquidity and creditworthiness of Treasuries in repurchase agreements (repos) to secure financing. In a repo transaction, Treasuries are used as collateral to borrow funds, which can then be employed to buy or sell cryptocurrency perpetual swaps. The repo market is highly liquid, allowing for favorable rates due to the high quality of Treasuries as collateral. This approach provides low-cost financing and flexibility, enabling quick access to funds.

The cost of funding through repos is influenced by U.S. Treasury yields. Repo rates are closely tied to short-term interest rates, which often track Treasury yields. Lower Treasury yields typically result in lower repo rates, reducing the cost of financing. Changes in the yield curve can impact repo rates; for example, a steepening yield curve might indicate rising short-term rates, increasing the cost of repo financing. Additionally, rising Treasury yields due to expectations of higher interest rates may signal tighter monetary policy, potentially affecting funding costs.

In cryptocurrency markets, funding rates are crucial, particularly in perpetual swaps. Positive funding rates, where long positions pay short positions, can increase the cost of holding long positions, while negative rates provide cost benefits. Differences between spot prices and perpetual swap prices, influenced by funding rates, create arbitrage opportunities that can be exploited using repo-financed capital. Balancing a portfolio between crypto derivatives and U.S. Treasuries involves managing risk and return. Treasuries provide stability and predictable yields, while cryptocurrencies offer higher returns but with greater volatility. Evaluating yield spreads between crypto yields and Treasury yields helps in determining asset allocations. A wider spread in favor of crypto might justify higher allocations despite the increased risk.

3.2 PERFORMANCE MEASUREMENT

Hedging strategies are essential for managing risks associated with interest rate movements and cryptocurrency volatility. Interest rate futures or swaps can hedge against adverse movements in Treasury yields that affect repo financing costs. Options and futures can hedge against crypto price volatility, balancing risk exposure between crypto and risk-free portfolio components. A diversified portfolio strategy might involve investing a portion in Treasuries to provide stable income and liquidity through repos while allocating funds to cryptocurrencies and perpetual swaps to capture higher returns and exploit arbitrage opportunities.

Yield management involves continuously monitoring the yield environment and adjusting allocations based on changes in Treasury and crypto yields. Dynamic rebalancing using machine learning or quantitative models can help respond to market conditions, yield spreads, and funding costs. Risk management through diversification across different cryptocurrencies and Treasury maturities is crucial, alongside leverage control to avoid margin calls in volatile markets. By leveraging Treasuries for repo financing, investors can engage in cryptocurrency perpetual swap trading with competitive funding rates. Managing the interaction between Treasury yields and cryptocurrency yields is vital for optimizing funding strategies and balancing the risk-reward profile of a mixed portfolio of crypto derivatives and risk-free assets. Implementing a dynamic, data-driven approach to portfolio management can help adapt to evolving market conditions and achieve optimal returns. performance measurement

Evaluating the performance of a balanced portfolio that includes cryptocurrencies, crypto derivatives, U.S. Treasuries, and Treasury derivatives involves several key metrics that help assess the risk, return, and overall efficiency of the portfolio. These metrics are essential for understanding how well the portfolio performs relative to its goals and market benchmarks.

Return metrics, such as total return and annualized return, provide a measure of the overall gain or loss of the portfolio over a specific period. Total return includes capital appreciation and income from interest or dividends, while the annualized return represents the geometric average annual return, making it easier to compare with other investments.

Risk metrics are crucial for understanding the volatility and downside potential of the portfolio. Volatility, measured by standard deviation, indicates the dispersion of returns around the mean, with higher volatility suggesting greater risk. Downside risk focuses on the potential for negative returns, capturing only the volatility of returns below a specified threshold or minimum acceptable return.

Risk-adjusted return metrics help evaluate how well the portfolio compensates for the risks taken. The Sharpe Ratio (Eq. 1) measures the portfolio's return relative to its risk by dividing the excess return (above the risk-free rate) by the standard deviation. A higher Sharpe Ratio indicates better risk-adjusted performance. The Sortino Ratio (Eq. 2), similar to the Sharpe Ratio, considers only downside risk and provides a clearer picture of how well the portfolio compensates for adverse movements.

$$Sharpe\ ratio = \frac{Portfolio\ Return - Risk\ Free\ Rate}{Volatility} \quad (1)$$

$$Sortino\ ratio = \frac{\text{Portfolio Return} - \text{Risk Free Rate}}{\text{Downside } Volatility} \quad (2)$$

Diversification metrics, such as the correlation matrix, analyze the correlation between different asset classes within the portfolio. Low or negative correlations indicate effective diversification, reducing overall portfolio risk. Beta measures the portfolio's sensitivity to market movements; a beta less than one suggests less volatility than the market, while a beta greater than one indicates more volatility.

Yield metrics are important for income-generating portfolios. Current yield reflects the income generated by the portfolio as a percentage of the current value, particularly relevant for assets like Treasuries. A yield spread compares the yield of the portfolio’s fixed-income components to benchmarks, such as the yield on Treasury bills or other government bonds, providing insights into the portfolio's income-generating efficiency.

Portfolio efficiency metrics, like the information ratio, compare the portfolio’s excess return over a benchmark to the tracking error, indicating the manager’s ability to generate alpha. Jensen’s Alpha measures the excess return of the portfolio over the expected return based on the portfolio’s beta and the market return, assessing the manager's skill in generating returns above the market.

Applying these metrics to portfolio analysis helps investors gain a comprehensive understanding of the performance of a balanced portfolio. For example, suppose a portfolio has an annualized return of 12%, with a volatility of 15%. The Sharpe Ratio is 0.8, and the Sortino Ratio is 1.2, indicating relatively good risk-adjusted performance with more emphasis on downside protection. If the correlation between cryptocurrencies and Treasuries is low, this suggests effective diversification, reducing overall risk. Additionally, if the yield on Treasuries in the portfolio is higher than comparable benchmarks, this indicates favorable yield positioning.

3.3 MODEL ASSUMPTIONS

We assume to have a fixed allocation among four different types of assets allowed in this portfolio. The target allocation for cryptocurrency, crypto derivatives, US treasury and US treasury derivatives in percentage are p1, p2, p3 and p4 respectively. We have a rebalance threshold of 0.10 as required by the business.

Among all the holdings in this portfolio, cryptocurrency perpetual swap has a funding interval every 8 hours. i.e. the funding rate payments are generally exchanged at regular intervals, commonly every 8 hours. This is standard across different types of cryptocurrency perpetual swaps, including Bitcoin, Ethereum, Ripple (XRP), Litecoin, and Chainlink. The common exchange time are: 04:00 UTC, 12:00 UTC, and 20:00 UTC. Therefore, our rebalance occurred 3 times a day during US treasury bond trading day.

Despite the many different type of cryptocurrencies in the market, our portfolio only hold the following five type to keep the problem simplified: Bitcoin, Ethereum, Ripple (XRP), Litecoin, and Chainlink.

3.4 DEEP Q-LEARNING

ALGORITHM 1. DEEP Q-LEARNING ALGORITHM

```
Initialize replay buffer
Initialize action-value function Q with random weights θ
Initialize target action-value function Q' with weights θ' = θ
Set contrastive learning network with shared layers for
representation learning
Initialize repo rate model
Define target allocations:
  target_crypto = p1
  target_crypto_derivatives = p2
  target_treasuries = p3
  target_treasury_derivatives = p4
Set rebalance threshold w = 0.10
for each episode do
  Initialize state s
  for each step, in episode do
    With probability ε select a random action a
    Otherwise select a = argmax_a' Q(s, a'; θ)

    Execute action a in the environment
    Observe reward r and next state s'
    Calculate current portfolio allocations, update:
      current_crypto
      current_crypto_derivatives
      current_treasuries
      current_treasury_derivatives

    Check if rebalancing is needed:
     if |current_crypto - target_crypto| > w
        or
|current_crypto_derivatives - target_crypto_derivatives| > w
        or
    |current_treasuries - target_treasuries| > w
        or
   |current_treasury_derivatives - target_treasury_derivatives| > w
then
        Perform rebalancing:
          Sell overweighted assets
          Buy underweighted assets
          Adjust repo positions accordingly

    Calculate financing cost:
      repo_cost = repo_rate * amount_financed

    Adjust reward for repo cost:
      r_adjusted = r - repo_cost

    Store transition (s, a, r_adjusted, s') in replay buffer

    Sample random batch of transitions (s_j, a_j, r_j, s'_j) from
replay buffer
    Compute Q-learning loss:
      L_Q = 1/N Σ_j [(r_j + γ * max_a' Q'(s'_j, a'; θ') - Q(s_j, a_j;
θ))^2]

    Perform contrastive learning:
      Sample pairs of states (s_j, s_k) from the replay buffer
      Compute contrastive loss:
        if s_j and s_k are similar (e.g., same market condition):
          L_C = ||f(s_j; θ) - f(s_k; θ)||^2  // Minimize distance
for similar pairs
        else:
          L_C = max(0, margin - ||f(s_j; θ) - f(s_k; θ)||)^2  //
Maximize distance for dissimilar pairs

    Combine losses:
      L_total = L_Q + λ * L_C

    Update network weights θ using ∇_θ L_total

    Periodically update target network:
      θ' = θ

    Set state s = s'
  end for
end for
```

The key hyperparameters for Deep Q-Learning play a crucial role in shaping the learning process and include several important factors. The learning rate (α) determines how quickly the neural network updates the Q-values when receiving new information, with lower values leading to smaller updates and higher values causing larger adjustments. The discount factor (γ) balances the importance of future versus immediate rewards, where a value closer to 1 prioritizes future rewards and a lower value emphasizes short-term gains. The exploration rate (ε) manages the trade-off between exploring new actions and exploiting known strategies, with higher values encouraging exploration and lower values favoring exploitation. This exploration rate often decays over time, controlled by the exploration decay rate, which reduces exploration as the model learns more about the environment.

In addition to these, the batch size defines how many past experiences are sampled from the replay buffer to update the Q-network during training. The replay buffer size determines how many past experiences the model stores and samples from, ensuring efficient learning. Another important hyperparameter is the target network update frequency, which dictates how often the target Q-network is updated with the weights of the online Q-network, providing stability in learning. The number of episodes and maximum steps per episode set the length of training, with episodes defining the total learning duration and the step limit ensuring the agent doesn’t get stuck in an infinite loop within an episode. Together, these hyperparameters shape the behavior and performance of the Deep Q-Learning model in optimizing decision-making.

Incorporating a rebalancing strategy into the Deep Q-Learning framework with repo funding involves creating a mechanism to adjust the portfolio allocations when they deviate from the target due to market movements. This strategy ensures that the portfolio remains aligned with the desired risk-return profile, even when using leverage through repos.

3.5 Contrastive Learning

The key for contrastive learning as mentioned in Section 2.3 is the reply buffer and Q-learning loss. The replay buffer stores experiences (state, action, reward, next state) to sample batches for training. This is crucial for stabilizing Q-learning by breaking the correlation between consecutive experiences. Q-Learning Loss (L_Q) is the standard loss function for Q-learning, which aims to minimize the difference between predicted Q-values and target Q-values calculated from rewards and next states.

State Pair Sampling is performed independent of trading every eight-hour period to be consistent with the cash flow exchange. States are sampled in pairs to determine whether they are similar (e.g., similar market conditions) or dissimilar. If bitcoin value increases in this period, we will create a contra-state pair (which has a negative outcome utility for market decrease).

The contrastive loss function *(L_C)* encourages the model to learn representations that are close for similar states and far apart for dissimilar states. This is typically done using distance metrics, such as Euclidean distance, and applying a margin to distinguish between similar and dissimilar pairs.

The total loss *(L_total)* combines the Q-learning loss and contrastive loss, where λ is a hyperparameter that balances the influence of contrastive learning. This helps the network to simultaneously learn effective Q-values and meaningful state representations.

Network Updates: The neural network is updated using the combined loss, which involves backpropagating the gradients and adjusting weights to minimize the total loss. A separate *target network*, with weights θ', is used to stabilize training. The target network's weights are periodically updated to match the primary network.

The contrastive learning step produces a learning set and an opposite contrastive set each time and is fed to the Q-network. This is a crucial step to add-in critical training condition compared to other approaches.

3.6 Portfolio Rebalance General Flow

The general flow structure for our system is as follows. We generated the contrastive training set (as shown in section 3.5) periodically and feed to the Q-learning network. The training set is supplied from trading records, as well as market/sentiment update. Not only each individual buy/sell action will become training set to Q network, but portfolio managers in this studied case also often conduct periodic review of their holdings and past trading activities. Successful (positive learning elements) and unsuccessful trades (contrastive learning elements) from the periodic review will also be formatted as training elements for Q-network.

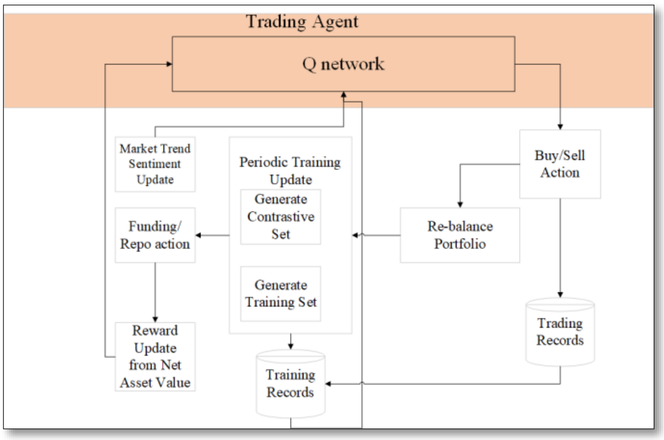

Figure 1. FLOWCHART FOR Q LEARNING UPDATE

Similar to [1], we define the net asset value (NAV) reward as in Eq. (3):

$$Reward = \frac{Changed\ composition - Original\ composition}{Original\ composition} \quad (3)$$

After each rebalance of the portfolio, updated training data with construction of contrastive set and market trend/sentiment update is feedback to Q-network. We are conducting periodic portfolio rebalancing, not the dynamic type introduced in [1].

3.7 Proximal Policy Optimization (PPO)

Reinforcement Learning (RL) is a type of machine learning where an agent learns to make decisions by interacting with an environment. In the context of trading strategies, such as predicting and trading cryptocurrency perpetual swaps, RL can be used to develop adaptive strategies that evolve based on changing market conditions. One popular RL algorithm is Proximal Policy Optimization (PPO). PPO is a type of policy gradient method used in reinforcement learning. It aims to optimize the policy (the strategy that dictates how actions are chosen based on states) in a way that balances exploration and exploitation while ensuring stable learning.

The key concept of PPO revolves around optimizing the policy to maximize cumulative rewards. The policy itself is a function that maps states to actions, guiding the agent's decision-making process. The advantage function plays a crucial role by measuring how much better or worse a particular action is compared to the average action in a given state. To ensure stable learning, PPO employs a clipped objective function, which prevents policy updates from deviating too significantly from the previous policy, thereby promoting controlled, incremental improvements.

When applying PPO to trading strategies, the first step is to define the trading environment. This involves creating an environment that accurately simulates the trading scenario, including various components such as the state space, action space, and reward function. The state space represents features of the market, such as price, volume, and funding rates, that provide the necessary inputs for decision-making. The action space encompasses the possible actions the trading agent can take, including buying, selling, or holding a position. Finally, the reward function is designed to measure trading performance, typically in terms of profit or loss, thus guiding the agent to optimize its strategy for maximum returns.

ALGORITHM 2. PROXIMAL POLICY OPTIMIZATION

```
Initialize actor network π with weights θ
Initialize critic network V with weights φ
Initialize repo rate model

Define target allocations:
    target_crypto = p1
    target_crypto_derivatives = p2
    target_treasuries = p3
    target_treasury_derivatives = p4

Set rebalance threshold = 0.10
Set clip parameter ε for PPO
Set discount factor γ
Set learning rates for actor and critic

for each episode do
    Initialize state s
    for each step in episode do
        Select action a based on policy π(a|s; θ)

        Execute action a in the environment
        Observe reward r and next state s'

        Calculate current portfolio allocations:
            current_crypto = current_value(crypto) / total_portfolio_value
            current_crypto_derivatives = current_value(crypto_derivatives) / total_portfolio_value
            current_treasuries = current_value(treasuries) / total_portfolio_value
            current_treasury_derivatives = current_value(treasury_derivatives) / total_portfolio_value

        Check if rebalancing is needed:
            if |current_crypto - target_crypto| > rebalance_threshold or
               |current_crypto_derivatives - target_crypto_derivatives| > rebalance_threshold or
               |current_treasuries - target_treasuries| > rebalance_threshold or
               |current_treasury_derivatives - target_treasury_derivatives| > rebalance_threshold then

                Perform rebalancing:
                    Sell overweighted assets
                    Buy underweighted assets
                    Adjust repo positions accordingly

        Calculate financing cost:
            repo_cost = repo_rate * amount_financed

        Adjust reward for repo cost:
            r_adjusted = r - repo_cost

        Store transition (s, a, r_adjusted, s') in memory
    end for
    Calculate advantages A for each transition:
        A(s, a) = Σ_t' [r_t' + γ * V(s_t'+1; φ) - V(s_t'; φ)]

    Update actor network π:
        Sample batch of transitions (s, a, A) from memory
        for each transition in batch do
            Compute ratio ρ:
                ρ = π(a|s; θ) / π_old(a|s; θ)

            Compute actor loss L_actor:
                L_actor = -min(ρ * A, clip(ρ, 1-ε, 1+ε) * A)

        Update θ using ∇_θ L_actor
    Update critic network V:
        Sample batch of transitions (s, r_adjusted) from memory
        Compute critic loss L_critic:
            L_critic = (V(s; φ) - r_adjusted)^2

        Update φ using ∇_φ L_critic

    Update policy parameters π_old = π

end for
```

First, the actor network (π) is initialized with specific weights (θ), while the critic network (V) is initialized with its respective weights (φ). A repo rate model is also initialized. Portfolio allocation targets are defined.

During each episode, the agent selects an action based on the policy, executes the action, and receives a reward. The portfolio allocations are then recalculated, and if any asset class deviates from its target allocation by more than the set threshold, a rebalancing action is triggered. Overweighted assets are sold, underweighted assets are purchased, and repo positions are adjusted accordingly. The repo cost is calculated and used to adjust the reward, which is stored in memory.

Advantages are calculated for each transition to determine the benefits of the actions compared to the estimated value. The actor network is updated using the probability ratio (ρ) between the new and old policies, and the critic network is updated with adjusted rewards to refine the value estimates. The old policy is periodically updated to match the new policy.

The key components of this approach revolve around network initialization, rebalancing strategy, repo funding, advantage calculation, and updates to the actor and critic networks. First, the actor network is tasked with determining the policy that guides actions, while the critic network evaluates the value of these actions. The rebalancing strategy compares the current portfolio allocations to target allocations and initiates rebalancing actions when deviations exceed a predetermined threshold. Repo funding is factored into the process by calculating the repo costs and adjusting rewards, accordingly, ensuring that financing costs are reflected during optimization. The advantage calculation is then used to evaluate how well an action performs relative to the estimated value. Finally, updates are made to both the actor and critic networks: the actor network is refined using a clipped objective to ensure stable learning, while the critic network is updated based on adjusted rewards to help the system learn to estimate long-term value.

This process ensures that the PPO framework dynamically adjusts to market movements, financing costs, and risk-return profiles while managing the balance between leverage and rebalancing strategies, ultimately keeping the portfolio aligned with investment objectives.

3.8 DIFFERENCE BETWEEN Q-LEARNING AND PPO

Deep Q-Learning and Proximal Policy Optimization (PPO) are two popular reinforcement learning algorithms with fundamental differences in their approach to decision-making and optimization, particularly in the context of managing a crypto-risk-free asset portfolio. Deep Q-Learning is a value-based method that focuses on learning a Q-value function, which estimates the expected return of taking a certain action in a given state and following the optimal policy thereafter. It is best suited for problems with discrete action spaces, as it requires calculating Q-values for each possible action. The agent selects actions based on these Q-values using an epsilon-greedy strategy to balance exploration and exploitation.

On the other hand, PPO is a policy-based method that directly learns a policy function, which maps states to actions by optimizing the expected return. It maintains a stochastic policy, allowing for exploration by sampling actions from a probability distribution. This makes PPO more flexible for complex decision-making problems as it can handle both continuous and discrete action spaces.

In terms of optimization and stability, Deep Q-Learning relies on the Bellman equation to iteratively update Q-values. This can lead to instability in certain environments due to the correlation between Q-value updates and policy changes, which is mitigated by using experience replay to improve stability and convergence. PPO, however, uses a clipped surrogate objective function that penalizes large deviations from the current policy, improving training stability. It often employs Generalized Advantage Estimation (GAE) to compute more accurate advantage estimates, leading to more efficient learning.

When considering sample efficiency and exploration, Deep Q-Learning can be less sample-efficient, especially in high-dimensional state spaces, because it relies heavily on exploring various state-action pairs to learn accurate Q-values. Its epsilon-greedy strategy for exploration can be less effective in environments with large or continuous action spaces. PPO, in contrast, tends to be more sample-efficient due to its policy-based nature, as it directly optimizes the policy using sampled trajectories. Its stochastic policy representation inherently encourages exploration, which can lead to better exploration in complex environments.

In the context of the crypto-risk-free portfolio problem, Deep Q-Learning may be suitable if the portfolio management decisions can be discretized into specific actions, such as discrete rebalancing percentages or binary decisions to buy or sell specific assets. However, it may struggle with the complexity and continuous nature of financial markets and portfolios, especially when considering continuous leverage adjustments and dynamic rebalancing. PPO, with its ability to handle continuous decision-making and more complex action spaces, is generally more suitable for this problem. It can adjust leverage, rebalance continuously, and optimize both risk and return in a dynamic environment. Its stability and capability to handle continuous action spaces make it a better fit for managing portfolios where precise adjustments are needed to optimize the portfolio's risk-return profile.

In summary, while both Deep Q-Learning and PPO can be applied to managing a crypto-risk-free portfolio, PPO is generally more effective due to its ability to handle continuous action spaces, improved sample efficiency, and robust policy optimization, making it more adept for dynamic portfolio management where decisions involve adjusting allocations and leverage continuously in response to changing market conditions.

4 RESULT DISCUSSION

The hypothesis for the study is: each portfolio can only buy any of the five cryptocurrencies and its derivatives; it is not allowed to buy/sell cryptocurrency outside of the five cryptocurrencies.

Portfolio performances are evaluated using *Sharpe ratio* and *Sortino ratio*. Sharpe ratio evaluates the portfolio's return relative to its risk by dividing the excess return (above the risk-free rate) by the standard deviation. A higher Sharpe ratio indicates better risk-adjusted performance. Sortino ratio is similar to the Sharpe ratio but considers only downside risk. It divides the excess return by the downside deviation, providing a clearer picture of how well the portfolio compensates for adverse movements. The detailed comparison of Sharpe ratio and Sortino ratio can be found in [18]. All risk-free rates are obtained from [19]. For cryptocurrency and its derivatives price are downloaded from Yahoo Finance.

We compare our solution with result applied to CNN model [12], and the FCNN model [12] introduced in the same paper. The combined LSTM models and reinforcement learning algorithms from [11] is also added into comparison, we denote it as LSTM-RL. Our training data mainly comes from 2020 and 2021, and we monitored the portfolio performance in 2022 and 2023 for comparison among methodologies that we are interested in this study. We use the parameter discussed in section 3.4 and 3.7 for our algorithm Q-Learning, Q-Learning-CL, PPO and PPO-CL, unless otherwise mentioned. The Deep-Q Learning and PPO models are implemented in Python using *PyTorch* and trained on an Intel i9 CPU with an RTX 4090 GPU. The training was conducted with a progressive learning rate adaptation strategy over 32 epochs, focusing on training the trading actions with market sentiment change.

Now we test our algorithm with CNN, DNN and LSTM-RLM for a portfolio of BTC, ETH, US treasury and their derivatives. Q-Learning-CL denotes the Q-Learning algorithm with contrastive learning. PPO-PL denotes the PPO algorithm with contrastive learning. The portfolio manager are permitted to execute both long and short orders for the portfolio.

In 2022, the bitcoin index had a negative return of -64.24% [20]. However, most strategies can still make some profit through short the current position through cryptocurrency derivatives. Table 1 show that PPO with contrastive learning can achieve a return of 40.21% through good prediction of the downward market trend. The LSTM reinforcement learning had the worst performance with a negative return of 7.14%. The result of CNN and DNN achieves 4.72% and 2.49% return of investment. The two deep learning methods: Q-Learning and PPO show promising results with 4.65% and 28.12% return over 2022.

TABLE 1. 2022 PORTFOLIO PERFORMANCE

	Cumulative Return	Sharpe Ratio	Sortino Ratio
CNN	4.72%	-1.70	-2.254
DNN	2.49%	-5.188	-6.008
LSTM-RL	-7.14%	-5.518	-6.186
Q-Learning	4.65%	-4.920	-5.592
Q-Learning-CL	10.97%	-3.361	-5.093
PPO	28.12%	-4.328	-4.042
PPO-CL	40.21%	-2.577	-3.443

Q-learning can perform well in environments where the portfolio adjustments are limited to discrete decisions, such as rebalancing between a few selected assets or choosing specific trading signals. In February and March 2023, where the market is facing the fear of economic crisis triggered by the collapse of Silicon Valley Bank and Signature Bank, PPO performs better than Q-learning. Q-Learning may also be more suitable in relatively stable market conditions where historical patterns and rewards can be relied upon to train the model. The decline in 2022 does not have a similar historical model to follow, therefore PPO performs better than Q-Learning.

In 2023, the bitcoin index had a positive return of 155.68% [20]. CNN performs best among all the algorithms we used for testing. When contrast learning is incorporated with Q-Learning and PPO, they are performing better than when no contrast learning. Q-learning with CL is having a portfolio return of 18.24% in 2023, where PPO-CL is having a return of 39.13% for 2023.

TABLE 2. 2023 PORTFOLIO PERFORMANCE

	Cumulative Return	Sharpe Ratio	Sortino Ratio
CNN	44.18%	10.041	22.778
DNN	1.37%	7.905	16.213
LSTM-RL	-0.11%	7.533	14.070
Q-Learning	0.64%	8.007	15.742
Q-Learning-CL	18.24%	8.656	17.846
PPO	32.99%	10.404	21.689
PPO-CL	39.13%	10.520	26.023

Q-Learning is best suited for simpler, discrete decision environments where trading actions are limited and predictable, leading to potentially better ROI in stable markets. PPO is more effective in dynamic, complex environments with continuous decision-making, leading to potentially better ROI and higher Sharpe ratios in volatile markets.

5 CONCLUSION AND FUTURE DIRECTION

In this paper, we explore the integration of contrastive learning with Deep Q-Learning and Proximal Policy Optimization (PPO) for managing a balanced portfolio consisting of five major cryptocurrencies—Bitcoin, Ethereum, Ripple (XRP), Litecoin, and Chainlink—and U.S. Treasuries. The combined approach demonstrated superior performance over traditional methods in both downward and upward cryptocurrency markets. In the bear market of 2022, the strategies proved effective in mitigating losses, while in the bull market of 2023, they capitalized on opportunities for significant gains. The results suggest that incorporating contrastive learning with reinforcement learning techniques enhances portfolio management across varying market conditions.

Throughout this study we don’t focus on the risk management perfective of a portfolio of cryptocurrency and risk-free-asset. Crypto derivatives are popular among cryptocurrency traders due to their flexibility, high leverage, and the ability to maintain positions without worrying about contract expirations. However, they also come with significant risks, including liquidation risk and the potential for significant losses due to high leverage.

There are many alternative metrics for measuring portfolio performance. The methodology may be adjusted if those alternative metrics are applied. For example, liquidity metrics, such as the liquidity ratio, assess the ease with which portfolio assets can be converted to cash without significant loss of value. This is important for portfolios with crypto derivatives and Treasuries, which may have different liquidity characteristics. The turnover rate indicates the frequency with which assets within the portfolio are bought and sold, with high turnover implying increased transaction costs and tax implications. We could revisit the portfolio management for cryptocurrency and US treasuries with the above metrics and re-evaluate all the methodologies to cater for those metrics.

ACKNOWLEDGMENTS

We would like to acknowledge Canoakbit Alliance Inc. to provide programming infrastructure for this research.

FUNDING

Not applicable.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

AUTHOR CONTRIBUTIONS

Conceptualization, ZL, BW and YC; methodology, ZL; software, ZL; validation, ZL, BW and YC; formal analysis, ZL, BW and YC; investigation, ZL, BW and YC; resources, YC; data curation, BW; writing—original draft preparation, ZL, BW and YC; writing—review and editing, ZL, BW and YC; visualization, ZL. All authors have read and agreed to the published version of the manuscript.

ABOUT THE AUTHORS

LI, Zichao

University of Waterloo, Canada.

WANG, Bingyang

Emory University, USA.

CHEN, Ying

MyMap AI, USA.

REFERENCES

[1] Q. Y. E. Lim, Q. Cao and C. Quek, "Dynamic portfolio rebalancing through reinforcement learning," Neural Computing and Applications, December 2021.

[2] E. Chen, M. Ma and Z. Nie, "Perpetual future contracts in centralized and decentralized exchanges: Mechanism and traders’ behavior," Electronic Markets, vol. 34, June 2024.

[3] Y.-S. Ren, C.-Q. Ma, X.-L. Kong, K. Baltas and Q. Zureigat, "Past, present, and future of the application of machine learning in cryptocurrency research," Research in International Business and Finance, vol. 63, p. 101799, December 2022.

[4] H. Sebastião and P. Godinho, "Forecasting and trading cryptocurrencies with machine learning under changing market conditions," Financial Innovation, vol. 7, January 2021.

[5] H. Ni, S. Meng, X. Geng, P. Li, Z. Li, X. Chen, X. Wang

and S. Zhang, "Time Series Modeling for Heart Rate Prediction: From ARIMA to Transformers," arXiv preprint arXiv:2406.12199, 2024.

[6] X. Li, J. Chang, T. Li, W. Fan, Y. Ma and H. Ni, "A Vehicle Classification Method Based on Machine Learning," Preprints, 2024.

[7] X. Li, Y. Yang, Y. Yuan, Y. Ma, Y. Huang and H. Ni, "Intelligent Vehicle Classification System Based on Deep Learning and Multi-Sensor Fusion," Preprints, July 2024.

[8] G. Lucarelli and M. Borrotti, "A deep Q-learning portfolio management framework for the cryptocurrency market," Neural Computing and Applications, vol. 32, September 2020.

[9] L. Weng, X. Sun, M. Xia, J. Liu and Y. Xu, "Portfolio trading system of digital currencies: A deep reinforcement learning with multidimensional attention gating mechanism," Neurocomputing, vol. 402, pp. 171-182, August 2020.

[10] Y. Li, P. Hu, Z. Liu, D. Peng, J. T. Zhou and X. Peng, "Contrastive Clustering," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 8547-8555, May 2021.

[11] Y. Li, S. Jiang, Y. Wei and S. Wang, "Take Bitcoin into your portfolio: a novel ensemble portfolio optimization framework for broad commodity assets," Financial Innovation, vol. 7, August 2021.

[12] Z. Jiang and J. Liang, "Cryptocurrency portfolio management with deep reinforcement learning," 2017 Intelligent Systems Conference (IntelliSys), September 2017.

[13] Z. Qi, D. Ma, J. Xu, A. Xiang and H. Qu, "Improved YOLOv5 Based on Attention Mechanism and FasterNet for Foreign Object Detection on Railway and Airway tracks," 2024. [Online]. Available: https://arxiv.org/abs/2403.08499.

[14] A. Xiang, B. Huang, X. Guo, H. Yang and T. Zheng, "A Neural Matrix Decomposition Recommender System Model based on the Multimodal Large Language Model," 2024. [Online]. Available: https://arxiv.org/abs/2407.08942.

[15] W. Zhu and T. Hu, "Twitter Sentiment analysis of covid vaccines," in 2021 5th International Conference on Artificial Intelligence and Virtual Reality (AIVR), 2021.

[16] Y. Wang, J. Zhao and Y. Lawryshyn, "GPT-Signal: Generative AI for Semi-automated Feature Engineering in the Alpha Research Process," in Proceedings of the Eighth Financial Technology and Natural Language Processing and the 1st Agent AI for Scenario Planning, Jeju, 2024.

[17] S. D’Amico and N. A. Pancost, "Special Repo Rates and the Cross-Section of Bond Prices: The Role of the Special Collateral Risk Premium," Review of Finance, vol. 26, October 2021.

[18] P. Srivastava and S. S. Mazhar, "Comparative Analysis of Sharpe and Sortino Ratio with reference to Top Ten Banking and Finance Sector Mutual Funds," International Journal of Management Studies, vol. V, p. 93, October 2018.

[19] U. D. o. t. Treasury, "Daily Treasury Par Yield Curve Rates," [Online]. Available: https://home.treasury.gov/resource-center/data-chart-center/interest-rates/. [Accessed 11 8 2024].

[20] "Historical performance of the Bitcoin index," Backtest by CURVO, [Online]. Available: https://curvo.eu/backtest/en/market-index/bitcoin. [Accessed 11 8 2024].