Joint Optimization of Order Allocation and Rack Selection in the “Parts-to-Picker” Picking System Considering Multiple Stations Workload Balance

Fang Wang 1, Yu Wang 1,∗ and Daofang Chang 2

1 Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China
2 Logistics Engineering College, Shanghai Maritime University, Shanghai 201306, China
∗ Correspondence: wangyu@shmtu.edu.cn

Abstract: E-commerce companies generate massive orders daily, and efficiently fulfilling them is a critical challenge. In the “parts-to-picker” order fulfillment system, the joint optimization of order allocation and rack selection is a crucial problem. Previous research has primarily focused on these two aspects separately and has yet to consider the issue of workload balancing across multiple picking stations, which can significantly impact picking efficiency. Therefore, this paper studies a joint optimization problem of order allocation and rack selection for a “parts-to-picker” order picking system with multiple picking stations to improve order picking efficiency and avoid uneven workload distribution. An integer programming model of order allocation and rack selection joint optimization is formulated to minimize the racks’ total moving distance and to balance the orders allocated to each picking station. The problem is decomposed into three sub-problems: order batching, batch allocation, and rack selection, and an improved simulated annealing (SA) algorithm is designed to solve the problem. Two workload comparing operators and two random operators are developed and introduced to the SA iterations. Random instances of different scales are generated for experiments. The algorithm solutions are compared with those generated by solving the IP model directly in a commercial solver, CPLEX, and applying the first-come-first-serve strategy (FCFS), respectively. The numerical results show that the proposed algorithm can generate order allocation and rack selection solutions much more efficiently, where the moving distances of the racks are effectively reduced and the workloads are balanced among the picking stations simultaneously. The model and algorithm proposed in this paper can provide a scientific decision-making basis for e-commerce companies to improve their picking efficiency.

Keywords: “parts-to-picker” picking system; order allocation; rack selection; workload balancing; improved simulated annealing algorithm

1. Introduction

In recent years, the proliferation of internet users and the growth of e-commerce platforms have made online shopping a significant mode of consumption. In 2021, the number of online shoppers in China was estimated at 842 million. The substantial increase in users presents a challenge for e-commerce warehouses, which must process an increasing volume of orders. The efficient processing of orders and the completion of picking activities have thus become a pressing issue for companies operating in this domain [1]. Order picking refers to extracting products from storage locations in a warehouse to fulfill customer orders [2]. The traditional “picker-to-parts” picking system entails workers continuously visiting racks, incurs non-productive costs and labor expenses, and is ill-equipped to meet the growing demand for order fulfillment [3]. With the advancement of Internet of Things technology and intelligent devices, the “parts-to-picker” order picking system has emerged to address these challenges, which becomes a crucial means for e-commerce warehouses to enhance order picking efficiency [4,5].
To improve the efficiency of order processing in distribution centers, major e-commerce companies have adopted “parts-to-picker” robotic mobile fulfillment systems (RMFS) with higher degrees of automation, such as Amazon’s Kiva system [6,7]. As depicted in Figure 1, RMFS consists of a series of mobile robots, movable racks, picking stations, other hardware systems, encompassing inbound docks, outbound docks, storage area, picking area, and storage and charging areas for the handling robots [8]. Goods are stored on movable racks in the warehouse and are transported by robots between picking stations and storage areas, significantly reducing the labor costs and workload of pickers and improving order picking efficiency [9].

![Figure 1. Warehouse layout of the robotic mobile fulfillment system.](image)

However, the “parts-to-picker” picking system has also brought challenges. The daily order amount for large warehouses with multiple picking stations can reach tens of thousands, with various products in each order. The stations work in parallel with a shared storage strategy, i.e., the same rack can store different products, and the same product can be stored on different racks. Thus, multiple rack selection options exist for the same order set. The critical decisions include selecting the optimal combination of racks, allocating orders to picking stations, and efficiently completing all orders with the shortest rack moving distance [10,11]. For rack selection, the ordered quantity of goods and the storage capacity of goods on the racks must be matched. Moreover, since the orders are too large to handle and some may require overlapped items, the orders can be batched before rack selection to improve the working efficiency. For order allocation, since multiple racks could provide each given product, it is necessary to consider the allocation of orders among multiple stations with the decisions on rack selection simultaneously. The joint optimization of the two decisions is more challenging, and efficient algorithms are required to avoid conflicts of racks and improve efficiency.

Our literature review found a lack of research on order allocation and rack selection in “parts-to-picker” picking systems that consider the parallel operation of multiple picking stations and the workload balancing across these stations.

Therefore, we consider a joint optimization for the order allocation and rack selection problem with multiple picking stations for the “parts-to-picker” mobile robot fulfillment system, taking the workload balance among the stations into account. An integer pro-
gramming model is formulated to minimize the workload imbalance rate and the total moving distance of the selected racks. We also designed an improved simulated annealing algorithm to solve the problem efficiently. In an order-picking system with multiple picking stations, the pickers work in parallel, so workload balance has to be considered during order allocation and rack selection. If the workload is balanced, then the overall completion time of the entire warehouse system could be completed, which will cause employee dissatisfaction or even affect the picking efficiency in the long term.

The innovation and contribution of this paper can be summarized as follows:

(1) A more comprehensive mathematical programming model is developed, simultaneously addressing order batching, batch allocation, and rack selection while considering constraints on rack storage capacity and workload balancing for picking stations.

(2) The paper considers the shared storage strategy of racks in the “parts-to-picker” system and considers order-rack similarity when batching orders.

(3) We propose an enhanced simulated annealing algorithm. Specifically, we have designed two workload comparison operators and two random exchange operators based on the objective. The former serves as a targeted domain action to facilitate a faster workload balance, while the latter aims to expand the solution space. By integrating the two operators, we perform iterative optimization of the initial solution to identify optimal solutions.

The remaining parts of this paper are organized as follows: Section 2 reviews the relevant literature. Section 3 provides a detailed description of the joint optimization problem and the integer programming model. Section 4 proposes the improved simulated annealing algorithm, which first introduces the generation of initial solutions and then designs neighborhood search operators to find better solutions iteratively. Section 5 verifies the effectiveness and efficiency of the proposed model and algorithm through extensive numerical experiments. Section 6 provides further conclusions and discussions.

2. Literature Review

In the current state of research, there is limited study on the joint optimization problem of order allocation and rack selection in the “parts-to-picker” system. Most studies focused on considering each decision-making aspect separately, making it challenging to obtain optimal solutions.

The problem of rack selection in the “parts-to-picker” mobile robot fulfillment system refers to selecting a collection of racks with the shortest total transportation distance or the minimum total number of racks based on the number of product types required by a batch of orders and the number of product types stored on the racks. Furthermore, this combination of racks must satisfy the picking requirements of the batch of orders. The rack selection problem is unique to the “parts-to-picker” system and is a relatively new combinatorial optimization problem that has received relatively little research attention. Li et al. [12] studied the rack selection problem for a given batch of orders by establishing a 0–1 linear programming model, aiming to minimize the round-trip time of shelves and proving its NP-hardness. They proposed a three-stage hybrid heuristic algorithm to solve this model. Wang Zheng [13] also assumed a known set of orders to be picked and studied the best rack selection problem by establishing an integer programming model and designing a simulated annealing algorithm with six local search operators to update the initial solution. Through various example scenarios, their findings demonstrated the superiority of their algorithm. Similarly, Zhang Tingting [14] established an integer programming model to minimize the number of times shelves are moved and designed a heuristic algorithm to solve it. The study proved that rack selection in e-commerce warehousing systems directly impacts picking efficiency and cost. However, these studies only consider single-picking platforms and do not consider similar picking scenarios for multiple-picking platforms commonly found in existing warehouses.

The allocation of orders in the “parts-to-picker” mobile robot fulfillment system refers to grouping or sorting orders to maximize picking efficiency while minimizing duplicated
resources and labor. About this, Wang Shanshan [15] proposed an improved genetic algorithm to solve an order batch optimization model for a shuttle storage system to minimize the total number of container outbounds. This approach resulted in a reduced number of outtakes and improved the overall picking efficiency of the system. Li Zhenping [16] conducted a similar study of order batching in a shuttle-based “parts-to-picker” picking system and analyzed the impact of parameters such as the capacity of the picking station and the similarity weighting coefficient on the batching results. The research showed that as the capacity of the picking station increased, the number of outbound containers decreased. Boysen et al. [17] studied the order allocation and rack selection problem in a single picking station and designed a simulated annealing decomposition program to address each sub-problem. Yang et al. [18] jointly optimized order allocation and rack selection in a single picking station in a robot movement performance system. Despite this, neither study accounted for the specific quantities of products on the racks, assuming their mere presence was sufficient to fulfill orders.

Furthermore, the joint optimization problem has also been explored by various researchers. Valle and Beasley [19] addressed the combined issue of order allocation, rack allocation, and rack ordering by assigning customer orders and mobile racks to pickers and determining the rack sequence within a single picking station. Wang et al. [20] investigated the problem of order picking in a robot movement performance system with multiple picking stations, intending to reduce the number of rack movements. This problem encompasses sub-problems such as order allocation, sorting, rack selection, and scheduling. It considers the specific consideration of inter-station rack selection and the queueing effect at each picking station, proposing a two-stage hybrid heuristic algorithm. Qin et al. [21] created two-stage A* and adaptive large neighborhood search algorithms. They used these algorithms to solve the problem of order allocation and path planning in multiple picking stations. This helps in efficiently scheduling and configuring e-commerce warehousing resources and provides practical decision-making guidance for e-commerce warehousing intelligence. Zhao Jinlong et al. [1] have developed an optimized order-picking model that considers delivery deadlines. They propose improved algorithms and rules to optimize order allocation, sorting decisions, and rack access orders. Despite the research on joint optimization problems, these studies have not considered the workload balance in a multi-picking station system, which does not reflect the warehouse layout.

In summary, the existing literature, both domestic and foreign, has laid a theoretical foundation for this paper. However, our investigation has revealed that there is currently a lack of research in the joint optimization of order allocation and rack selection with workload balancing for multiple picking stations in the “parts-to-picker” picking system. Specifically, previous research has yet to address the challenge of determining order batches for picking stations and the appropriate set of racks while ensuring workload balance and minimizing the distance traveled by the racks. Thus, there is a need for further research on the joint optimization problem of order allocation and rack selection that considers workload balancing while minimizing the rack movement distance in order to improve the efficiency and workflow of the “parts-to-picker” picking system.

3. Problem Description and Formulation

3.1. Problem Description

The order picking process of the “parts-to-picker” system is shown in Figure 2 below. For the customer order set obtained from the e-commerce information system, the system first performs order allocation, including order batching and batch allocation, and establishes a correspondence between orders and picking stations. Workload balancing is essential as we consider multiple picking stations in the “parts-to-picker” picking system. Specifically, previous research has yet to address the challenge of determining order batches for picking stations and the appropriate set of racks while ensuring workload balance and minimizing the distance traveled by the racks. Thus, there is a need for further research on the joint optimization problem of order allocation and rack selection that considers workload balancing while minimizing the rack movement distance in order to improve the efficiency and workflow of the “parts-to-picker” picking system.
racks to be picked need to be transported by robots to the corresponding picking stations and then returned to their storage locations after picking, the walking distance of the robot is a critical factor that affects the efficiency of order picking. The robot’s walking path includes three parts: (1) from the current location to the target rack; (2) transporting the rack to the corresponding picking station; and (3) returning the rack to its storage location after picking. Among them, (1) is the distance traveled without a load, while (2) and (3) are the distances traveled with a load. It can be seen that the distance traveled with a load (i.e., the rack moving distance) accounts for the central part of the total distance; therefore, minimizing the rack moving distance is crucial for improving order picking efficiency [10].

![Illustration of order allocation and rack selection.](image)

**Figure 2.** Illustration of order allocation and rack selection.

Furthermore, considering the limited capacity of the picking station, workers can handle up to \( l \) orders simultaneously, and the number of orders in each batch can be at most \( l \). The problem is to batch orders, allocate them to picking stations, determine the combination of racks, and minimize the total rack moving distance while balancing the workload of each picking station.

### 3.2. Assumptions

In this study based on a rack-moving mobile robot picking system, the following assumptions are made:

- The location of the racks in the warehouse is fixed, and the racks return to their initial storage position after completing the picking task;
- All racks have the same specification attributes;
- All picking stations have the same capacity;
- The rack inventory can meet the order demand;
- Orders in the same batch can be processed simultaneously without considering the order or rack order within the same batch.

### 3.3. Integer Programming Model

The descriptions of relevant parameters and variables can be found in Tables 1 and 2.
Table 1. Description of symbols and parameters.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(J)</td>
<td>Order Set, (J = {j</td>
</tr>
<tr>
<td>(R)</td>
<td>Rack Set, (R = {r</td>
</tr>
<tr>
<td>(B)</td>
<td>Batch Set, (B = {b</td>
</tr>
<tr>
<td>(P)</td>
<td>Picking Station Set, (P = {p</td>
</tr>
<tr>
<td>(K)</td>
<td>Commodity Category Set, (K = {k</td>
</tr>
<tr>
<td>(I)</td>
<td>The maximum number of orders a batch can accommodate, which is also the maximum order capacity a picking station can handle simultaneously</td>
</tr>
<tr>
<td>(W_p)</td>
<td>The workload of picking station (p)</td>
</tr>
<tr>
<td>(a_j)</td>
<td>The number of SKUs contained in order (j)</td>
</tr>
<tr>
<td>(a_{jk})</td>
<td>The number of SKUs (k) required by order (j)</td>
</tr>
<tr>
<td>(d_{rk})</td>
<td>The storage capacity of SKU (k) on rack (r)</td>
</tr>
<tr>
<td>(l_{rp})</td>
<td>The distance from rack (r) to picking station (p)</td>
</tr>
<tr>
<td>(c_1)</td>
<td>Representative unit product picking cost</td>
</tr>
<tr>
<td>(c_2)</td>
<td>Representative unit rack movement cost</td>
</tr>
<tr>
<td>(q_{br})</td>
<td>The number of SKUs (k) provided by rack (r) for batch (b)</td>
</tr>
</tbody>
</table>

Table 2. Decision variables.

<table>
<thead>
<tr>
<th>Decision Variables</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x_{jb})</td>
<td>0–1 Variables: 0 indicates that order (j) is not assigned to batch (b); 1 indicates that order (j) is assigned to batch (b)</td>
</tr>
<tr>
<td>(y_{bp})</td>
<td>0–1 Variables: 0 indicates that batch (b) is not assigned to picking station (p); 1 indicates that batch (b) is assigned to picking station (p)</td>
</tr>
<tr>
<td>(z_{br})</td>
<td>0–1 Variables: 0 indicates that batch (b) does not select rack (r) as its provider of SKU; 1 indicates that batch (b) selects rack (r) as its provider of SKU</td>
</tr>
<tr>
<td>(w_{jp})</td>
<td>0–1 Variables: 0 indicates that order (j) is not processed by picking station (p); 1 indicates that order (j) is processed by picking station (p)</td>
</tr>
<tr>
<td>(v_{brp})</td>
<td>0–1 Variables: 0 indicates that rack (r) is not assigned to picking station (p) when it processes batch (b); 1 indicates that rack (r) is assigned to picking station (p) when it processes batch (b)</td>
</tr>
</tbody>
</table>

Objective Function:

(a) Minimize the level of imbalance in the workload of the picking stations, i.e., minimize the difference between the maximum and minimum workloads of the picking stations:

\[
F_1 = \min (\max W_p - \min W_p)
\]  

(b) To minimize the rack movement distance:

\[
F_2 = 2 \min \sum_{p=1}^{s} \sum_{b=1}^{u} \sum_{r=1}^{n} v_{br}^p l_{rp}
\]  

Since the uneven workload allocation increases the waiting time for workers, in order to reduce the waiting time, the unit product picking cost is introduced, transforming the objective function (a) into a penalty cost incurred due to the uneven workload allocation. At the same time, the unit distance movement cost is introduced to transform the objective function (b) into the rack movement cost. Additionally, two weight coefficients are introduced. Thus, the multi-objective function is transformed into a single objective function, as follows:

\[
\min F = \omega_1 c_1 F_1 + \omega_2 c_2 F_2
\]  

Constraints:
(a) Each order can only be assigned to one batch:

$$\sum_{b=1}^{u} x_{jb} = 1 \forall j$$

(4)

(b) The total number of orders in each batch cannot exceed its maximum capacity:

$$\sum_{j=1}^{m} x_{jb} \leq I \forall b$$

(5)

(c) A batch can only be assigned to one picking station:

$$\sum_{p=1}^{s} y_{bp} = 1 \forall b$$

(6)

(d) Each order is assigned to the picking station to which its batch is posted:

$$\omega_{jp} = x_{jb} \times y_{bp} \forall j b p$$

(7)

(e) Each batch selects the racks to go to the corresponding picking station:

$$v_{br}^{p} = y_{bp} \times z_{br} \forall b r p$$

(8)

(f) The number of products offered by the racks cannot exceed their storage quantity:

$$\sum_{b=1}^{m} q_{br}^{k} \leq d_{rk} \forall b r k$$

(9)

(g) The product requirements of each batch order are met:

$$\sum_{j=1}^{m} x_{jb} a_{jk} = \sum_{r=1}^{n} q_{br}^{k} \forall b k$$

(10)

(h) The range of values for the decision variables:

$$x_{jb}, y_{bp}, z_{br}, \omega_{jp}, v_{br}^{p}, q_{br}^{k} \in \{0, 1\}$$

(11)

$$q_{br}^{k} \in N$$

(12)

4. Improved Simulated Annealing Algorithm

The joint optimization problem of order allocation and rack selection is an NP-hard problem. This paper proposes an improved simulated annealing algorithm to rapidly and effectively solve this problem. The algorithm first generates an initial solution: an improved seed algorithm is used to batch the orders based on the order-rack similarity; a random strategy is used to allocate the batches to the picking stations; and a greedy algorithm based on the rack cost–performance ratio generates the rack selection scheme. Subsequently, two types of neighborhood search operators are designed to generate new neighboring solutions: workload comparison operators and random operators. If the neighboring solution is better than the current solution, it replaces it; otherwise, it is accepted with a probability of exp\((\frac{-\Delta f}{T})\).

4.1. Order-Rack Similarity

Order similarity is a numerical value that measures the similarity of the order’s attributes, and allocating orders with high similarity to the same batch can reduce repeated
picking and improve picking efficiency [22,23]. In traditional “picker-to-parts” order picking systems, the order batching is mainly based on the similarity of order items, which refers to the ratio of the number of standard products contained in two orders to the total number of products. Orders usually contain various products in the “parts-to-picker” system considered in this paper. The same rack can store different products, and different products on the same rack can be picked simultaneously. Therefore, considering order item similarity in this system is difficult to obtain an effective order batching result. This paper proposes order-rack similarity, which is similar to order-item similarity [17] and is defined as \( R_{jr} = (R_{j1r}, R_{j2r}, \ldots, R_{jr})^T \).

\[
R_{jr} = \begin{cases} 
1, & \text{The existence of a certain product in order } j \text{ being stored on rack } r \\
0, & \text{Otherwise} 
\end{cases}
\]

The similarity between orders \( j \) and \( j' \) in terms of their respective racks is calculated as follows:

\[
S_{jj'} = \frac{R_{jr}^T R_{jr}}{R_{jr}^T R_{jr} + R_{jr'}^T R_{jr'} - R_{jr}^T R_{jr'}}
\]

**Example:** Suppose four products are stored in racks \( R_1 = \{a, a, a, b, b\} \), \( R_2 = \{a, a, c, c\} \), \( R_3 = \{b, b, b, d\} \), and \( R_4 = \{b, c, c, d\} \), and there are three orders to be picked, \( J_1 = \{b\} \), \( J_2 = \{a, a\} \), and \( J_3 = \{c, d\} \). Then, for \( R_{1r} = (1, 0, 1, 1)^T \), \( R_{2r} = (1, 1, 0, 0)^T \), and \( R_{3r} = (0, 1, 1, 1)^T \), the rack similarity between order one and order two is \( S_{12} = 0.25 \), between order two and order three is \( S_{23} = 0.25 \), and between order one and order three is \( S_{13} = 0.5 \).

### 4.2. Generation of Initial Solutions

This section presents a heuristic algorithm to generate the initial solution for the improved simulated annealing algorithm, consisting of the following three parts:

**Part 1: Generation of Initial Batch Result Based on the Similarity of Order Racks**

Step 1: Calculate the ratio of the number of orders \( m \) and the maximum number of orders that a batch can accommodate \( I \) and round it up to \( \mu = \lceil \frac{m}{I} \rceil \);

Step 2: Calculate the similarity \( S \) between any two orders, sort the order similarity values \( S \) in descending order, and select the first \( \mu - \text{order} \) pairs with high similarity values as seed order pairs;

Step 3: Calculate the average similarity value of the remaining orders with each seed order pair and add it to the seed order pair with the maximum average similarity value until reaching the batch capacity limit or there are no orders to be added;

Step 4: Output the batch result \( B \).

**Part 2: Random Generation of Order Batch Allocation**

Step 1: Calculate the picking workload for each batch \( W_b = a_j x_{jb} \) based on the order batch result, wherein \( a_j = \sum_{k=1}^{i} a_{jk} \).

Step 2: Sort the batches in descending order of picking workload.

Step 3: Allocate the batches to picking stations sequentially.

Step 4: Output the order batch allocation result \( X_0 \) and calculate the difference between the maximum and minimum picking station workload \( F_1 \).

**Part 3: Greedy Algorithm Based on Rack Cost-effectiveness for Generating Initial Rack Selection Solution**

We propose a rack cost-effectiveness method to determine which racks are more cost-effective. This method prioritizes racks with higher cost-effectiveness because each rack’s moving distance is different.
Step 1: Initialize $R = (1, 2, \cdots, n)$, $Y = \emptyset$; use vector $Q_b = (a_{b1}, a_{b2}, \cdots, a_{bk})^T$ ($b = 1, 2, \cdots, u$) to represent the vector of the demand for all goods in batch $b$, where $a_{bk}$ represents the amount of goods $k$ contained in batch $b$.

Step 2: Remove redundant racks. Set $b = 1$, and calculate $E = \sum_{k=1}^{t} a_{bk} d_{rk}$ for any rack $r = \{1, 2, \cdots, n\}$. If $E = 0$, remove the rack from $R$, effectively reducing the search space and improving solution speed; otherwise, go to step 3.

Step 3: Calculate the cost–performance of the rack. Define the cost–performance of the rack as a rate:

$$rate(r) = \frac{\sum_{k=1}^{t} q_{br} l_{rp}}{l_{rp}}$$

where: $q_{br} = \min\{a_{bk}, d_{rk}\}$ represents the number of products $k$ that rack $r$ provides for this batch, and $l_{rp}$ represents the distance of rack $r$ from picking station $p$.

Step 4: Sort the racks in descending order of cost–performance.

Step 5: Add the rack with the highest cost–performance to set $Y$ and remove it from $R$, updating $Q_b$.

Step 6: If $Q_b = 0$, go to step 6; otherwise, go to step 2.

Step 7: Set $b = b + 1$. If $b > u$, go to step 7; otherwise, go to step 1.

Step 8: Calculation ends; output the rack set $Y$ corresponding to each batch and calculate the distance $F_2$ required to move all racks for picking all batches.

4.3. Improved Simulated Annealing Algorithm

In order to search for better solutions to the problem at hand, two types of neighborhood search operators have been designed: workload comparison operators (Operators 1 and 2) and random exchange operators (Operators 3 and 4). Workload comparison operators steer the perturbation towards a balanced workload, serving as a targeted neighborhood action that leads to a faster workload balance. Random exchange operators are advantageous for expanding the solution space. Combining these two types of operators improves computational efficiency and increases the probability of obtaining the optimal global solution.

Revised Neighborhood Operation:

Operator 1: Select the order batch with the enormous workload from the picking station with the enormous workload and add it to the picking station with the minor workload, ensuring that the new solution satisfies the batch capacity constraint of the picking station.

Operator 2: Randomly select an order batch $b$ from the picking station with the most considerable workload and exchange it with an order batch $b'$ from the picking station with the minor workload where $W_h > W_{hp}$.

Operator 3: Randomly select an order and add it or exchange it with another batch, ensuring that the batch capacity constraint is satisfied after addition.

Operator 4: Randomly select two picking stations; select one batch from each and exchange the batches.

The specific steps of the algorithm are as follows:

Step 1: Initialization. The order allocation solution $X = X_0$ obtained in the above stage, the total cost $F = f(X)$ of the objective function, the initial temperature $T = T_0$, and the number of iterations $I$;

Step 2: Under the current temperature $T$, randomly select a neighborhood search operator to perturb the current solution, obtain a new solution $X'$, and calculate the corresponding total cost $f(X')$;

Step 3: Calculate the incremental cost $\Delta f = f(X') - f(X)$. If $\Delta f < 0$, accept $X'$ as the new current solution, $X = X'$, $F = f(X')$, $I = I + 1$; otherwise, accept the current solution with a probability of $\exp(-\frac{\Delta f}{T})$; that is, randomly generate a random number
Parameters for the different-sized cases are shown in Table 3. The improved simulated annealing algorithm proposed in this paper was compared with the CPLEX solver and the commonly used first-come-first-served (FCFS) strategy in practice, respectively. The initial temperature \( T \) of the improved SA was set to 500; the decay coefficient \( \alpha \) was set to 0.98; the number of iterations \( L \) was set to 100; and the final temperature \( T_{\text{min}} \) was set to 0.01. The results were finally compared with and without considering the balancing factor.

5. Experimental Results and Discussions

In order to verify the validity of the proposed model and algorithm, the case in the literature [20] was adopted, and simulations were conducted using small-scale \([10 \times 15]\), medium-scale \([10 \times 30]\), and large-scale \([20 \times 30]\) warehouse grid maps as shown in Figure 3. The distances were measured in grid units, and the distance between the racks and the picking stations was their Manhattan distance. The racks, picking stations, and robots are the same size and occupy one grid, and the robot walking step length is 1. The parameters for the different-sized cases are shown in Table 3. The improved simulated annealing algorithm proposed in this paper was compared with the CPLEX solver and the commonly used first-come-first-served (FCFS) strategy in practice, respectively. The initial temperature \( T \) of the improved SA was set to 500; the decay coefficient \( \alpha \) was set to 0.98; the number of iterations \( L \) was set to 100; and the final temperature \( T_{\text{min}} \) was set to 0.01. The results were finally compared with and without considering the balancing factor.

The methods were implemented using the Python programming language, and the experimental environment was Windows 10, Intel(R) Core(TM) i7-10875H CPU @ 2.30 GHz with 16.0 GB RAM.
Table 3. Setting of Different-Sized Instances.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Small-Scale Instances</th>
<th>Medium-Scale Instances</th>
<th>Large-Scale Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of picking stations ($s$)</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Number of racks ($n$)</td>
<td>50</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>SKUs</td>
<td>100</td>
<td>100/200</td>
<td>100/200</td>
</tr>
<tr>
<td>Batch capacity ($W_p$)</td>
<td>5</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Unit product picking cost ($c_1$)</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit rack movement cost ($c_2$)</td>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>weight coefficients</td>
<td>$\omega_1 = \omega_2 = 0.5$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.1. Comparison with CPLEX Results

The CPLEX solver and the improved simulated annealing algorithm proposed in this paper were used to solve small-scale cases. As shown in Table 4, when the number of orders is within 20, CPLEX can solve the problem within 1.1 s and find the optimal solution. However, as the number of orders increases, the computation time increases exponentially. When the number of orders is 40, the computation time reaches 7159.8 s, and CPLEX cannot solve the problem when the number of orders is 50. Compared to CPLEX’s solution time, the improved simulated annealing algorithm proposed in this paper has a significant advantage, with an average computational time of 138.3 s. Compared to CPLEX’s solution, the error rate of the improved simulated annealing algorithm proposed in this paper does not exceed 10% with an average error rate of 6.2%. When the number of orders is 30, the solution quality is the best with only a 1.9% error.

Table 4. Resolution Results of Small-Sized Instance.

<table>
<thead>
<tr>
<th>Scale of Example</th>
<th>CPLEX</th>
<th>ISA ¹</th>
<th>Gap (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td>Number of Orders</td>
<td>$F$</td>
<td>Running Time per Second (s)</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>3.6</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>20</td>
<td>3.7</td>
<td>1.1</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>5.1</td>
<td>203.3</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>7.1</td>
<td>7159.8</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Average</td>
<td>1841.2</td>
<td>138.3</td>
<td>6.2</td>
</tr>
</tbody>
</table>

¹ ISA: Improved Simulated Annealing Algorithm.

5.2. Comparison with FCFS Results

Due to difficulties in generating feasible solutions with CPLEX, our algorithm and the commonly used first-come-first-served (FCFS) strategy in the warehouse were applied to solve medium- and large-scale instances. In order to more clearly demonstrate their respective objectives for workload balancing and rack movement distance, a comparison was made between the two regarding their workload balancing level and rack distance. The results are shown in Tables 5 and 6.

Tables 5 and 6 demonstrate that the proposed algorithm significantly improves the workload balance. Compared to warehouses commonly used “first-come-first-served” allocation strategy, the proposed method reduces the rack-moving distance. With average reductions of 94.4% and 86.2%, the maximum workload imbalance can be reduced by 100%. The proposed method also results in average reductions of 47% and 47.2% in the maximum rack-moving distance, with a maximum reduction of 56.4%.” When the number of racks in the warehouse increases to 200, the total number of products to 200, and the number of orders to 400, the algorithm’s runtime remains acceptable.
Table 5. Resolution Results of Medium-Sized Instance.

<table>
<thead>
<tr>
<th>Example</th>
<th>SKUs</th>
<th>Number of Orders</th>
<th>£F1£</th>
<th>£F2£</th>
<th>£F1£</th>
<th>£F2£</th>
<th>Running Time per Second (s)</th>
<th>£F1£ Decline Rate</th>
<th>£F2£ Decline Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>100</td>
<td>50</td>
<td>18</td>
<td>264</td>
<td>2</td>
<td>120</td>
<td>210.2</td>
<td>88.9</td>
<td>44.4</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>100</td>
<td>27</td>
<td>530</td>
<td>1</td>
<td>238</td>
<td>343.5</td>
<td>96.3</td>
<td>55.1</td>
</tr>
<tr>
<td>9</td>
<td>150</td>
<td>26</td>
<td>868</td>
<td>0</td>
<td>406</td>
<td>519.3</td>
<td>666.0</td>
<td>91.0</td>
<td>50.2</td>
</tr>
<tr>
<td>10</td>
<td>200</td>
<td>33</td>
<td>1152</td>
<td>3</td>
<td>574</td>
<td>666.0</td>
<td>91.0</td>
<td>50.2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>100</td>
<td>50</td>
<td>25</td>
<td>396</td>
<td>3</td>
<td>268</td>
<td>220.8</td>
<td>88.0</td>
<td>32.3</td>
</tr>
<tr>
<td>12</td>
<td>100</td>
<td>100</td>
<td>29</td>
<td>916</td>
<td>0</td>
<td>460</td>
<td>410.4</td>
<td>100.0</td>
<td>49.8</td>
</tr>
<tr>
<td>13</td>
<td>150</td>
<td>18</td>
<td>1322</td>
<td>1</td>
<td>656</td>
<td>549.0</td>
<td>94.4</td>
<td>50.4</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>200</td>
<td>31</td>
<td>1958</td>
<td>1</td>
<td>1166</td>
<td>864.0</td>
<td>96.8</td>
<td>40.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Average</td>
<td>472.9</td>
<td>94.4</td>
</tr>
</tbody>
</table>

Table 6. Resolution Results of Large-Sized Instance.

<table>
<thead>
<tr>
<th>Example</th>
<th>SKUs</th>
<th>Number of Orders</th>
<th>£F1£</th>
<th>£F2£</th>
<th>£F1£</th>
<th>£F2£</th>
<th>Running Time per Second (s)</th>
<th>£F1£ Decline Rate</th>
<th>£F2£ Decline Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>100</td>
<td>100</td>
<td>32</td>
<td>448</td>
<td>1</td>
<td>226</td>
<td>663.0</td>
<td>96.9</td>
<td>49.6</td>
</tr>
<tr>
<td>16</td>
<td>150</td>
<td>150</td>
<td>24</td>
<td>778</td>
<td>1</td>
<td>386</td>
<td>1037.1</td>
<td>95.8</td>
<td>50.4</td>
</tr>
<tr>
<td>17</td>
<td>200</td>
<td>200</td>
<td>41</td>
<td>1010</td>
<td>1</td>
<td>594</td>
<td>1360.2</td>
<td>97.6</td>
<td>41.2</td>
</tr>
<tr>
<td>18</td>
<td>300</td>
<td>300</td>
<td>24</td>
<td>1390</td>
<td>3</td>
<td>846</td>
<td>1803.0</td>
<td>87.5</td>
<td>39.1</td>
</tr>
<tr>
<td>19</td>
<td>400</td>
<td>400</td>
<td>23</td>
<td>2128</td>
<td>2</td>
<td>1128</td>
<td>2160.0</td>
<td>91.3</td>
<td>47.0</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>100</td>
<td>23</td>
<td>816</td>
<td>4</td>
<td>432</td>
<td>767.7</td>
<td>82.6</td>
<td>47.1</td>
</tr>
<tr>
<td>21</td>
<td>150</td>
<td>150</td>
<td>22</td>
<td>1384</td>
<td>16</td>
<td>604</td>
<td>1172.2</td>
<td>27.3</td>
<td>56.4</td>
</tr>
<tr>
<td>22</td>
<td>200</td>
<td>200</td>
<td>27</td>
<td>1836</td>
<td>0</td>
<td>888</td>
<td>1584.3</td>
<td>100.0</td>
<td>51.6</td>
</tr>
<tr>
<td>23</td>
<td>300</td>
<td>300</td>
<td>18</td>
<td>2440</td>
<td>0</td>
<td>1382</td>
<td>2357.7</td>
<td>100.0</td>
<td>43.4</td>
</tr>
<tr>
<td>24</td>
<td>400</td>
<td>400</td>
<td>47</td>
<td>3848</td>
<td>8</td>
<td>2082</td>
<td>2374.2</td>
<td>83.0</td>
<td>45.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Average</td>
<td>1422.4</td>
<td>86.2</td>
</tr>
</tbody>
</table>

Furthermore, we observe that when the total number of Stock Keeping Units (SKUs) is 100, the running time is relatively short, and the reductions in the levels of unbalanced workload and rack distance are relatively high. On the other hand, when the total number of SKUs is 200, the running time is relatively long, and the reductions in the levels of unbalanced workload and rack distance are relatively low. The primary reason for this phenomenon is that when the total number of SKUs in the warehouse increases, the possibility of having the same SKUs in different orders decreases, which unavoidably requires more racks and results in a heavier picking workload. However, the proposed improved simulated annealing algorithm still exhibits excellent performance.

5.3. The Impact of Workload Balancing Factors

Table 7 compares six randomly selected cases from the set of instances to demonstrate the importance of workload balancing. The cases compare the picking station workload with and without considering the workload balancing objective to obtain the minimum rack moving distance. For the 6 cases, the solutions obtained by considering the workload balancing objective showed an average decrease of 92.27% in the imbalance of the picking station workload compared to the solutions obtained without considering this objective, with reductions of 83.3%, 94.1%, 100%, 98.1%, 84.2%, and 93.9%, respectively. On the other hand, the obtained rack moving distance increased by an average of 6.35%, with increases of 3.4%, 1.7%, 3.0%, 8%, 19.5%, and 2.5%, respectively. Figures 4 and 5 compare the maximum
workload difference and rack moving distance, respectively, between considering and not considering the balancing.

**Table 7.** The effect of balancing factors on the results.

<table>
<thead>
<tr>
<th>Example</th>
<th>Ignoring Workload Balance</th>
<th>Considering Workload Balance</th>
<th>Comparison of Enhancements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_1$</td>
<td>$F_2$</td>
<td>$F_1$ Decline Rate/$%$</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>116</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>238</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>27</td>
<td>394</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>53</td>
<td>550</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>19</td>
<td>708</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>33</td>
<td>1100</td>
<td>2</td>
</tr>
</tbody>
</table>

Average 92.3 −6.4

**Figure 4.** Comparing the maximum effort difference between considering equalization and not considering.

**Figure 5.** Comparison of rack travel distance between considering equilibrium and not considering.
Therefore, considering both factors, the improved simulated annealing algorithm proposed in this paper can effectively balance the order workload distribution while avoiding a significant increase in the rack moving distance.

5.4. Sensitivity Analysis of Weight Coefficients of Objective Functions

Due to the direct impact of weighting coefficients $\omega_1$ and $\omega_2$ on the solution of the simulated annealing algorithm in the objective function, this section selects a specific example from small-scale, medium-scale, and large-scale computations to conduct a sensitivity analysis of the maximum difference in workload between different picking stations and the total rack movement distance under varying $\omega_1$ and $\omega_2$. Figure 6 depicts the variation of the maximum workload difference as the weighting coefficients change. As the value of $\omega_1$ increases, the workload imbalance between different sorting stations gradually becomes balanced. It can also be observed that the balancing speed is faster when $0 \leq \omega_1 \leq 0.5$ and slower when $0.5 \leq \omega_1 \leq 1$. Figure 7 illustrates the variation of the rack movement distance as the weighting coefficients change. As the value of $\omega_1$ increases, the optimal rack movement distance gradually increases. The trends of the two objective functions are opposite and cannot be optimized simultaneously. Therefore, in practice, the weighting coefficients of the two objective functions should be determined based on specific circumstances to achieve satisfactory optimization results.

![Figure 6](image1.png)

**Figure 6.** The relationship between the maximum workload difference and the weight coefficient.

![Figure 7](image2.png)

**Figure 7.** The relationship between the distance of the selected racks’ movement and the weight coefficient.
6. Conclusions

This study explores the joint optimization of order allocation and rack selection in a “parts-to-picker” picking system, considering the picking stations’ balancing factor. A mixed-integer programming model is developed to minimize workload imbalance and the distance to selected racks while addressing three sub-problems: order batching, batch allocation, and rack selection. The correspondence between multiple orders, picking stations, and racks are established in the system. An improved simulated annealing algorithm is designed, which includes two workload balancing operators and two random exchange operators to improve its efficiency. The former balances the workload quickly, while the latter helps expand the solution space. Combining the two can increase the probability of obtaining the optimal global solution, essential for multi-picking station environments where workload balancing is crucial while minimizing rack movement distance.

In addition, the study considers the characteristics of the “parts-to-picker” system when solving the simulated annealing algorithm’s initial solution. A more scenario-specific order-rack similarity is defined to obtain more effective order-batching results. Furthermore, rack cost-effectiveness is proposed, prioritizing the selection of closer racks that can provide more of the required items. This approach can provide a good initial solution for the simulated annealing algorithm, enabling it to find the optimal solution quickly.

Finally, numerical experiments on different scales demonstrate the proposed algorithm’s effectiveness. The algorithm can provide an effective solution for order allocation and rack selection in e-commerce warehouses, reducing labor costs, improving picking efficiency, and having significant implications for warehouse management and operation. The findings can provide practical guidance and reference for future applications.

Moreover, as e-commerce warehouses face many random events in actual scenarios, such as cancellations and urgent orders, future research can explore the joint optimization of order allocation and rack selection that considers workload balancing under uncertain conditions. Additionally, introducing more decision factors, such as path planning and robot task allocation, can make the algorithm more applicable to the actual operating scenario of the “parts-to-picker” system.

Author Contributions: Conceptualization, F.W. and Y.W.; methodology, F.W. and D.C.; investigation, Y.W. and D.C.; writing—original draft preparation, F.W.; writing—review and editing, Y.W. and D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was sponsored by Shanghai Sailing Program (21YFI1416400).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.