Evolutionary Multitasking via 显式自动编码 Explicit Autoencoding

Liang Feng, Lei Zhou, Jinghui Zhong, Abhishek Gupta, Yew-Soon Ong, Kay-Chen Tan, and A. K. Qin, *Senior Member, IEEE*

***Abstract*—Evolutionary multitasking (EMT) is an emerging research topic in the field of evolutionary computation. In contrast to the traditional single-task evolutionary search, EMT conducts evolutionary search on multiple tasks simultaneously. It aims to improve convergence characteristics across multiple optimization problems at once by seamlessly transferring knowledge among them. Due to the efficacy of EMT, it has attracted lots of research attentions and several EMT algorithms have been proposed in the literature. However, existing EMT algorithms are usually based on a common mode of knowledge transfer in the form of *implicit genetic transfer* through chromosomal crossover. This mode cannot make use of multiple biases embedded in different evolutionary search operators, which could give better search performance when properly harnessed. Keeping this in mind, this paper proposes an EMT algorithm with *explicit genetic transfer* across tasks, namely EMT via autoencoding, which allows the incorporation of multiple search mechanisms with different biases in the EMT paradigm. To confirm the efficacy of the proposed EMT algorithm with explicit autoencoding, comprehensive empirical studies have been conducted on both the single- and multi-objective multitask optimization problems.**

***Index Terms*—Autoencoder, evolutionary optimization, knowledge transfer, multitask optimization.**

Manuscript received October 8, 2017; revised January 22, 2018, April 18, 2018; accepted June 5, 2018. Date of publication July 2, 2018; date of current version June 6, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61603064 and Grant 61602181, in part by the Frontier Interdisciplinary Research Fund for the Central Universities under Grant 106112017CDJQJ188828, in part by the Chong-Qing Application Foundation and Research in Cutting Edge Technologies under Grant cstc2017jcyjAX0319, in part by the City University of Hong Kong Research Fund under Grant 7200543, in part by the Program for Guangdong Introducing Innovative and Enterpreneurial Teams under Grant 2017ZT07X183, and in part by the Data Science and Artificial Intelligence Research Centre and the School of Computer Science and Engineering at Nanyang Technological University. This paper was recommended by Associate Editor G. G. Yen. *(Corresponding author: Liang Feng.)*

L. Feng and L. Zhou are with the College of Computer Science, Chongqing University, Chongqing 400044, China (e-mail: liangf@cqu.edu.cn; stone_zhou@cqu.edu.cn).

J. Zhong is with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China (e-mail: jinghuizhong@gmail.com).

A. Gupta and Y.-S. Ong are with the School of Science and Computer Engineering, Nanyang Technological University, Singapore (e-mail: asysong@ntu.edu.sg; abhishekg@ntu.edu.sg).

K.-C. Tan is with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: kaytan@cityu.edu.hk).

A. K. Qin is with the Department of Computer Science and Software Engineering, Swinburne University of Technology, Melbourne, VIC 3122, Australia (e-mail: kqin@swin.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2018.2845361

I. Introduction

Evolutionary algorithms (EAs) are adaptive search approaches that take inspirations from the principles of natural selection and genetics [1]. The algorithm starts with a population of individuals that undergo reproduction and mutation to produce a generation of offspring. The procedure executes iteratively and terminates when a predefined condition is satisfied. Due to the strong search capability and ease of use, in the last decades, EAs enjoyed significant successes in obtaining optimal or near-optimal solutions on a plethora of complex real-world optimization problems, including continuous optimization [2]–[5], combinatorial optimization [6]–[8], constrained optimization [9]–[13], etc.

Recently, inspired by the remarkable ability of the human brain which is able to perform multiple tasks with apparent simultaneity, an intriguing evolutionary search paradigm, namely evolutionary multitasking (EMT), has been proposed in the realm of evolutionary computation [14]. In contrast to traditional EA search which solves a single task in a single run, EMT conducts evolutionary search concurrently on multiple search spaces corresponding to different tasks or optimization problems, each possessing a unique function landscape [15]. By exploiting the latent synergies between distinct (but possibly similar) optimization problems,[1] superior search performances of EMT in terms of solution quality and convergence speed have been verified on a set of continuous, discrete, and the mixtures of continuous and combinatorial tasks [14]. Due to the efficacy demonstrated by EMT, increasing research interests in developing EMT algorithms for solving complex real-world problems have emerged in the literature. In particular, Gupta *et al.* [19] presented a realization of the EMT paradigm within the domain of multiobjective optimization (MOO). Zhou *et al.* [20] proposed a permutation-based unified representation and a *split*-based decoding operator for conducting EMT on the NP-hard capacitated vehicle routing problems. More recently, Ting *et al.* explored the resource allocation mechanism for reallocating fitness evaluations on offsprings of different tasks and proposed a framework, namely evolution of biocoenosis through symbiosis, for many-tasking optimization problem in [21] and [22], respectively. In [23], toward positive knowledge sharing across tasks, Bali *et al.* proposed a linearized domain adaptation strategy that transforms the

[1]To quantify the similarity between problems, different criteria could be used with respect to different perspectives, such as the problem landscape synergy [16], problem solution correlation [17], [18], etc.

search space of a simple task to the search space similar to its constitutive complex task. Moreover, Tang *et al.* [24] introduced an EMT algorithm for training multiple extreme learning machine with different number of hidden neurons for classification problems.

Despite the success enjoyed by the EMT search paradigm, it is worth noting that most of the current EMT algorithms are designed based on the *unified solution representation*, and the knowledge sharing across tasks for problem-solving is realized by the *implicit genetic transfer* in chromosomal crossover. In this manner, a common evolution mechanism (i.e., crossover and mutation) is usually designed for solving multiple tasks. However, today, it is well known that different optimization problems always have unique properties, which may require different evolution mechanisms with distinct search bias for efficient problem-solving [25], [26]. In the literature, various dedicated evolutionary solvers with unique search bias have been introduced. For example, the traditional one-point crossover has a very strong positional bias, but no distributional bias. The two-point crossover, on the other hand, can reduce the positional bias without introducing any distributional bias [27]. Generally, a search operator which is particularly useful for some classes of problems, is quite poor for others [28]. Therefore, toward enhanced EMT performance, it is desirable to incorporate multiple search mechanisms with problem-specific bias, while at the same time allow knowledge transfer across tasks in the EMT paradigm. Taking this cue, in this paper, we embark a study on the exploration of knowledge sharing across tasks in EMT via *explicit genetic transfer*. To the best of our knowledge, there is no existing study conducting EMT in an explicit manner with the incorporation of multiple search mechanisms for solving different tasks.

In particular, first of all, instead of designing a *unified solution representation* to represent multiple task domains concurrently, here we propose to use independent solution representation for each task,[2] respectively. Based on this, separate evolutionary solvers with different biases are then employed for solving different tasks simultaneously. Next, we introduce a denoising autoencoder for conducting knowledge transfer, in the form of problem solutions, across tasks explicitly. As the derived denoising autoencoder holds a closed-form solution, it will not bring much additional computational burden in the evolutionary search process. Further, since multiple dedicated evolutionary solvers can be incorporated in the proposed EMT for multiple tasks, more diverse and useful traits could be found and transferred along the search process when compared to the previous single solver EMT search paradigm. To verify the efficacy of the proposed EMT with explicit autoencoding, comprehensive empirical studies conducted on both the single objective [17] and multiobjective [18] multitask optimization benchmarks are presented.

The rest of this paper is organized as follows. A brief introduction of the concept of evolutionary multitask optimization is introduced in Section II. The general workflow of existing EMT search paradigm which employs a single solver for solving multiple tasks that conducts knowledge sharing through *implicit genetic transfer*, is also presented in this section. Further, Section III presents the proposed EMT paradigm with *explicit genetic transfer* via autoencoding, including the theoretical derivation of the single layer denoising autoencoder and the detailed design of the EMT search paradigm for single- and multi-objective tasks, respectively. Section IV provides the comprehensive empirical studies on the commonly used single objective as well as multiobjective EMT benchmarks. Lastly, the concluding remarks of this paper are discussed in Section V.

[2]We do not use unified representation for all the tasks as in [14]. Each task uses its own solution representation. For example, suppose we are given two tasks, i.e., one is 50-D Ackley function, while the other is 100-D Sphere function. Then the real coded solutions with 100-D and 50-D are used for Ackley and Sphere, respectively.

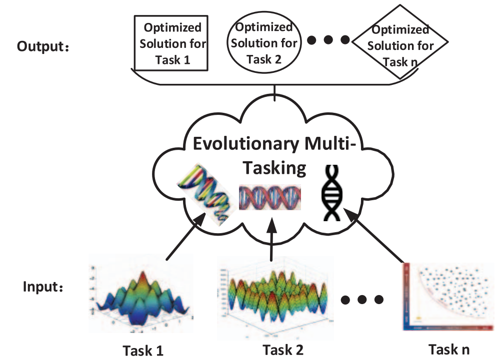


Fig. 1. Illustration of evolutionary multitask optimization.

II. PRELIMINARY

In this section, we first present a brief introduction of the concept of the evolutionary multitask optimization. Next, an introduction of existing EMT search paradigm which uses a single evolutionary solver with unified solution representation for solving multiple tasks is also provided.

A. Evolutionary Multitask Optimization

Generally, optimization problems can be classified into two groups: 1) single-objective optimization (SOO), where every point in the search space maps to a single scalar objective value or 2) MOO, where every point in the search space maps to a vector-valued objective function. To solve the given optimization problems, in the literature, most of the existing optimization methods, particularly EAs, are designed in a single-tasking manner. In particular, these algorithms solve only one optimization problem, i.e., either single-objective or multiobjective problem, in a single run. However, as it is well known that problems seldom exist in isolation [29] and knowledge learned from solving a problem can be deployed to enhance the optimization on related problems [30], the design of knowledge sharing toward enhanced optimization is desired in the context of evolutionary optimization.

Recently, evolutionary multitask optimization has been proposed to conduct evolutionary search on multiple optimization problems (or tasks) simultaneously. Inspired by human's cognitive ability to multitasking, the EMT aims to improve convergence characteristics across multiple optimization problems by seamlessly transferring knowledge between them. In particular, as depicted in Fig. 1, the input of the EMT consists of multiple optimization tasks, in which each could be either an SOO or MOO problem. The EMT paradigm optimizes all the tasks at the same time, thus the output of EMT contains the optimized solutions for each task, respectively. In contrast to the evolutionary single-task optimization, EMT optimizes multiple tasks simultaneously with the automatical exploitation and transfer of latent synergies between distinct (but possibly similar) optimization problems while the search progresses online, which could eventually lead to enhanced problem-solving on all the tasks.

B. Evolutionary Multitasking via Implicit Genetic Transfer

Currently, the particular mode of knowledge transfer in existing designs of EMT algorithms is that of *implicit genetic transfer* through chromosomal crossover [14], [23]. In particular, considering the situation wherein K optimization tasks are to be performed, a population of individuals in a *unified representation space* has to be defined. To compare the individuals in the population for multitasking, the following properties for every individual are usually defined in existing EMT approaches.

1) *Factorial Cost:* The factorial cost f_p of an individual p denotes its fitness or objective value on a particular task T_i. For K tasks, there will be a vector with length K, in which each dimension gives the fitness of p on the corresponding task.
2) *Factorial Rank:* The factorial rank r_p simply denotes the index of individual p in the list of population members sorted in ascending order with respect to their factorial costs on one specific task.
3) *Scalar Fitness:* The scalar fitness φ_p of an individual p is defined based on its best rank over all tasks, which is given by $\varphi_p = [1/(\min_{j\in\{1,\ldots,K\}} r_p^j)]$.
4) *Skill Factor:* The skill factor τ_p of individual p denotes the task, amongst all other tasks in MFO, on which p is most effective, i.e., $\tau_p = \operatorname{argmin}\{r_p^j\}$, where $j \in \{1, \ldots, K\}$.

With the properties described above, solutions comparison can be carried out in a straightforward manner based on the *scalar fitness*. For example, solution s_a is considered to dominate s_b in multitasking sense simply if $\varphi_a > \varphi_b$ [14]. The *skill factor* serves as the indicator that gives the task on which the current solution has best performance. Further, the general work-flow of existing EMT algorithms can be summarized as follows.

Step 1: Generate an initial population with NP individuals using a unified representation.
Step 2: Evaluate each individual on all the tasks by calculating its factorial cost f_p, factorial rank r_p, scalar fitness φ_p, and skill factor τ_p.
Step 3: Apply genetic operators, i.e., assortative mating, on the current population to generate an offspring population.
Step 4: Evaluate offspring individuals on selected tasks based on vertical cultural transmission.
Step 5: Update the scalar fitness φ_p and skill factor τ_p of individuals in both parent and offspring population.
Step 6: Select the fittest NP individuals based on the scalar fitness from both parent and offspring population to survive for the next generation.
Step 7: If the stopping criteria are not met, then repeat steps 3 to 6.

As can be observed, in this EMT framework, since the optimization of multiple tasks is based on a single population, a common genetic operator is designed for solving different tasks. The *implicit genetic transfer* across tasks is usually realized via this genetic operator. However, as aforementioned, since different tasks typically posses different problem properties, it is desirable to have multiple genetic operators with different biases for the tasks toward enhanced problem-solving in EMT. Keeping this in mind, in this paper, we present a new EMT algorithm with *explicit genetic transfer* for multitask optimization, which allows different genetic operators to be used for different tasks. For more details of the framework of exiting EMT algorithms, interested readers can refer to [14] and [15].

III. Proposed Evolutionary Multitasking via Explicit Autoencoding

In this section, the design of the proposed EMT paradigm with explicit genetic transfer across tasks is presented. In particular, we first give the theoretical derivation of the single layer denoising autoencoder, which servers as the key component for conducting knowledge transfer in the form of problem solutions across tasks. Next, the proposed EMT with autoencoding for both SOO and MOO is detailed.

A. Denoising Autoencoder

In recent years, deep learning has been proposed in the literature to discover intricate structure in large data sets [31]. It has been successfully applied for solving many challenging learning problems, such as image classification [32], natural language processing [33], transfer learning [34], [35], handwriting recognition [36], etc.

An *autoencoder* is the basic building block of deep learning networks that attempts to reproduce its input, i.e., the target output is equal to the input itself [37]. Typically, the autoencoder consists of an input layer, a hidden layer, and an output layer. Given the input vector $\mathbf{x} \in [0, 1]^d$, the encoding process converts the input samples to obtain hidden-layer representation $\mathbf{y} \in [0, 1]^{d'}$ via the mapping $\mathbf{y} = s(\mathbf{W}\mathbf{x} + \mathbf{b})$, where $\mathbf{W}$ and $\mathbf{b}$ represent the weight and bias between the input and the hidden layer, respectively, and s is the sigmoid activation function, i.e., $s(\mathbf{x}) = [1/(1 + e^{(-\mathbf{x})})]$. The decoding process is to reproject the encoded representation to a reconstructed vector $\mathbf{z} \in [0, 1]^d$ in the original signal space, where $\mathbf{z} = s(\mathbf{W}'\mathbf{y}+\mathbf{b}')$, such that $\mathbf{z} \approx \mathbf{x}$. The autoencoder parameters, i.e., $\mathbf{W}$, $\mathbf{W}'$, $\mathbf{b}$, $\mathbf{b}'$ are optimized to minimize the average reconstruction error as shown by

$$\min_{\mathbf{W},\mathbf{W}',\mathbf{b},\mathbf{b}'} \frac{1}{n}\sum_{i=1}^{n} L(\mathbf{x}_i, \mathbf{z}_i) \tag{1}$$

where n denotes the number of data instances, and L is a loss function, such as the square loss, Kullback–Leibler divergence, etc.

Denoising autoencoder is a stochastic version of autoencoder, which tries to encode the input in a way that can undo the effect of a stochastic corruption process applied to it [37]. In doing so, the learner must capture the structure of the input distribution in order to reduce the effect of the corruption process. Thus, the learned hidden representation is usually used as a new robust representation in many learning applications. However, it is worth noting that this hidden representation is also a connection between the corrupted inputs $\tilde{\mathbf{x}}$ and the repaired "clean" input $\mathbf{x}$. By configuring the input and output layer to represent two problem domains, the hidden representation provides a possibility for conducting knowledge transfer across domains [35]. Inspired from this, in this paper, we propose to conduct the EMT with *explicit genetic transfer* via denosing autoencoder. By treating the solutions of one problem domain as the corrupted version of the solutions of another problem domain, the connection between these two domains could be constructed by the denoising process. In particular, let $\mathbf{P}$ and $\mathbf{Q}$ represent the set of solutions uniformly and independently sampled from the search space of two different optimization problems $\mathbf{OP}_1$ and $\mathbf{OP}_2$, respectively, i.e., $\mathbf{P} = \{\mathbf{p}_1 \ldots, \mathbf{p}_N\}$ and $\mathbf{Q} = \{\mathbf{q}_1 \ldots, \mathbf{q}_N\}$, where N denotes the number of solutions in each set.[3] The connection from $\mathbf{OP}_1$ to $\mathbf{OP}_2$ can be naturally built through a denoising autoencoder by using $\mathbf{P}$ as the input and $\mathbf{Q}$ as the output accordingly.

Further, for simplicity, here we reconstruct the corrupted inputs with a single level mapping $\mathbf{M}$: $\mathcal{R}^d \to \mathcal{R}^d$ (d is the space dimension), that minimizes the squared reconstruction loss, which is given by

$$\mathcal{L}_{sq}(\mathbf{M}) = \frac{1}{2N}\sum_{i=1}^{N}\|\mathbf{p}_i - \mathbf{M}\mathbf{q}_i\|^2. \tag{2}$$

To simplify the notation, we assume a constant feature is added to the input, i.e., $\mathbf{p}_i = [\mathbf{p}_i; 1]$ and $\mathbf{q}_i = [\mathbf{q}_i; 1]$, and an appropriate bias is incorporated within the mapping $\mathbf{M} = [\mathbf{M}, \mathbf{b}]$. The loss in (2) is then reduced to the matrix form

$$\mathcal{L}_{sq}(\mathbf{M}) = \frac{1}{2N}\mathrm{tr}\big[(\mathbf{Q} - \mathbf{M}\mathbf{P})^{\mathrm{T}}(\mathbf{Q} - \mathbf{M}\mathbf{P})\big] \tag{3}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operation of a matrix. The solution of (3) can be expressed as the well-known closed-form solution for ordinary least squares [38], which is given by

$$\mathbf{M} = \left(\mathbf{Q}\mathbf{P}^{\mathrm{T}}\right)\left(\mathbf{P}\mathbf{P}^{\mathrm{T}}\right)^{-1}. \tag{4}$$

As can be observed, the learned $\mathbf{M}$ serves as the mapping from problem $\mathbf{OP}_1$ to $\mathbf{OP}_2$. The transfer of useful genetic solutions from $\mathbf{OP}_1$ to $\mathbf{OP}_2$ can thus be conducted simply by the multiplication operation with $\mathbf{M}$.

In what follows, the specific designs of the EMT with this derived single-layer denoising autoencoder for single- and multi-objective multitasking will be presented.

[3]Note that $\mathbf{p}$ and $\mathbf{q}$ may have different dimensions, and we pad $\mathbf{p}$ or $\mathbf{q}$ with zeros to make both problems be of equal dimensionality.

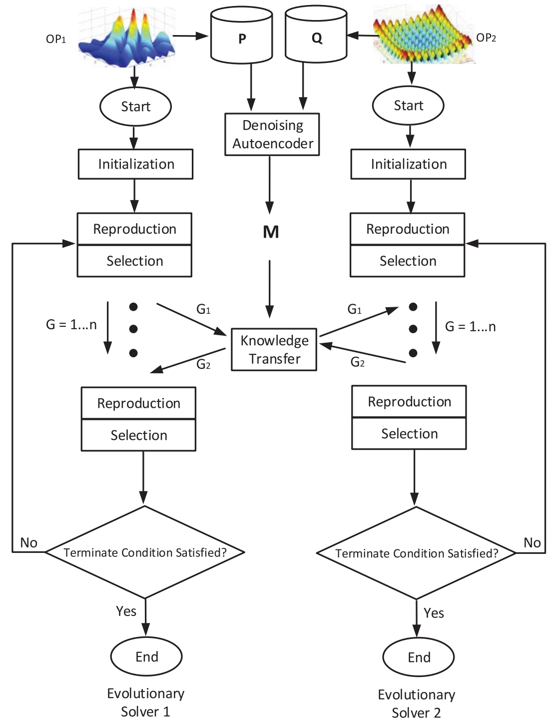


Fig. 2. Workflow of the proposed EMT with explicit genetic transfer.

B. Proposed EMT Paradigm

Fig. 2 gives the workflow of the proposed EMT paradigm with *explicit genetic transfer*. As can be observed, to optimize two optimization tasks, i.e., $\mathbf{OP}_1$ and $\mathbf{OP}_2$, concurrently, in contrast to existing EMT paradigm as discussed in Section II-B, two separate evolutionary solvers with independent population have been first employed. Next, to built the connection across tasks, two sets of problem solutions which are uniformly and independently sampled from the search space of $\mathbf{OP}_1$ and $\mathbf{OP}_2$, are then used as the input and output for the denoising autoencoder to obtain the corresponding task mapping $\mathbf{M}$ via (4). Further, the *explicit genetic transfer* happens along the searches of the two evolutionary solvers that progress online. In particular, good solutions found by any solver along its search can be simply transferred and injected into the population of the other through the multiplication operation with the learned task mapping $\mathbf{M}$. The populations of both solvers with the injected solutions will undergo the natural selection and reproduction until a certain stop criterion is satisfied. Note that, the proposed EMT paradigm is different from the cooperative co-evolution in the literature. Although both of them contains multiple evolutionary solvers, the proposed EMT is tackling multiple tasks with different objectives, while the cooperative co-evolution is solving a single task with a common objective [39].

1) Learning of Task Mapping: For any two optimization tasks, e.g., $\mathbf{OP}_1$ and $\mathbf{OP}_2$, two types of task mappings are required to conduct the genetic solution transfer across tasks, i.e., one is from $\mathbf{OP}_1$ to $\mathbf{OP}_2$, and the other is from $\mathbf{OP}_2$ to $\mathbf{OP}_1$. The pseudo code of learning the task mapping from $\mathbf{OP}_1$ to $\mathbf{OP}_2$ is outlined in Algorithm 1. As can be observed, two sets (i.e., $\mathbf{P}$ and $\mathbf{Q}$) with equal number of solutions are first

Algorithm 1: Pseudo Code of the Learning of Task Mapping From $\mathbf{OP}_1$ to $\mathbf{OP}_2$

Input : $\mathbf{OP}_1$ and $\mathbf{OP}_2$: two optimization problems with given objective functions and variable range definitions.
Output: $\mathbf{M}_{12}$: Learned task mapping from $\mathbf{OP}_1$ to $\mathbf{OP}_2$.

Begin
Sample two sets (with size N) of solutions, denoted by $\mathbf{P}$ and $\mathbf{Q}$, uniformly and independently based on the given variable range definition of $\mathbf{OP}_1$ and $\mathbf{OP}_2$, respectively.
Sort the solutions of $\mathbf{P}$ and $\mathbf{Q}$ in either ascending or descending order with respect to their objective values.
Configure the sorted $\mathbf{P}$ and $\mathbf{Q}$ as the input and output of the denoising autoencoder derived in Section III-A, respectively.
Obtain $\mathbf{M}_{12}$ through Eq. 4.
End

uniformly and independently sampled from the search space of $\mathbf{OP}_1$ and $\mathbf{OP}_2$, respectively. Next, the solutions in $\mathbf{P}$ and $\mathbf{Q}$ are sorted based on their objective values in either ascending or descending order. In particular, for SOO, the solutions are directly evaluated with the given single objective function. For MOO, the sorting process will be performed on each objective function of the MOO problem, which will result in multiple sorted solution sets. Further, by configuring the sorted $\mathbf{P}$ and $\mathbf{Q}$ as the input and output of the denoising autoencoder, the mapping $\mathbf{M}_{12}$ from $\mathbf{OP}_1$ to $\mathbf{OP}_2$ can be obtained through solving (4). Please note that, in the case that the two tasks have different dimensions, we will pad the corresponding solutions with zeros to make both problems be of equal dimensionality. Further, as multiple sorted $\mathbf{P}$ and $\mathbf{Q}$ are available in the context of MOO, there will be multiple mappings from $\mathbf{OP}_1$ to $\mathbf{OP}_2$ for MOO. For instance, as illustrated in Fig. 3, suppose there are 2 objectives in $\mathbf{OP}_1$ and $\mathbf{OP}_2$, respectively, there will be 4 task mappings from $\mathbf{OP}_1$ and $\mathbf{OP}_2$, i.e., $\mathbf{M}_{12}^{11}$, $\mathbf{M}_{12}^{12}$, $\mathbf{M}_{12}^{21}$, and $\mathbf{M}_{12}^{22}$. The subscript of $\mathbf{M}$ denotes the direction of the mapping, while the superscript of $\mathbf{M}$ gives the objective function used to sort the sampled solution in each task, e.g., $\mathbf{M}_{12}^{21}$ denotes the task mapping from $\mathbf{OP}_1$ and $\mathbf{OP}_2$ that uses the second function in $\mathbf{OP}_1$ and the first function in $\mathbf{OP}_2$ for sorting $\mathbf{P}$ and $\mathbf{Q}$. Note that the sorting procedure induces high ordinal correlation across objective function values of the task datasets, thereby facilitating positive genetic transfers through the learned mapping M as the search progresses [44], respectively.

On the other hand, by configuring $\mathbf{Q}$ and $\mathbf{P}$ as the input and output of the denoising autoencoder, and remaining the other three steps unchanged, the mapping $\mathbf{M}_{21}$ from $\mathbf{OP}_2$ to $\mathbf{OP}_1$ can be obtained accordingly.

2) Explicit Genetic Transfer Across Tasks: Besides the learning of mapping across tasks, how to conduct the explicit genetic transfer in the context of SOO and MOO is another key issue in the proposed EMT paradigm. In particular, when the transfer occurs and what to be transferred across tasks have to be specifically designed.

For SOO, as depicted in Fig. 2, the knowledge transfer across tasks occurs with a fixed generation interval along the search. In the proposed EMT, for simplicity, we configure $G_1 = G_2 = 10$, and conduct the transfer across tasks in every ten generations. However, without loss of generality, other sophisticated design of transfer frequency can also be applied. Further, toward a positive transfer, we choose the best S solutions with respect to the objective value in each population and transfer them across tasks through the multiplication of the learned task mappings $\mathbf{M}$s.

For MOO, the interval for knowledge transfer is configured as the same as that in SOO, i.e., $G_1 = G_2 = 10$. Next, to conduct genetic transfer across MOO problems, due to the existence of multiple objective functions, we propose to transfer the best S solutions of one solver to the other with respect to a randomly selected objective function. Further, as discussed in Section III-B1, multiple task mappings $\mathbf{M}$s shall be learned due to the use of different objective functions for sorting. The task mapping corresponds to the objective function which is used for selecting the best S solutions, will then be used for the solution transfer across tasks. For instance, given two MOO problems $\mathbf{OP}_1$ and $\mathbf{OP}_2$ and each has two objective functions, according to Section III-B1, we will have eight task mappings, i.e., four mappings from $\mathbf{OP}_1$ to $\mathbf{OP}_2$, and four from $\mathbf{OP}_2$ to $\mathbf{OP}_1$. The corresponding explicit genetic transfer across these two tasks will then be executed as outlined in Algorithm 2.

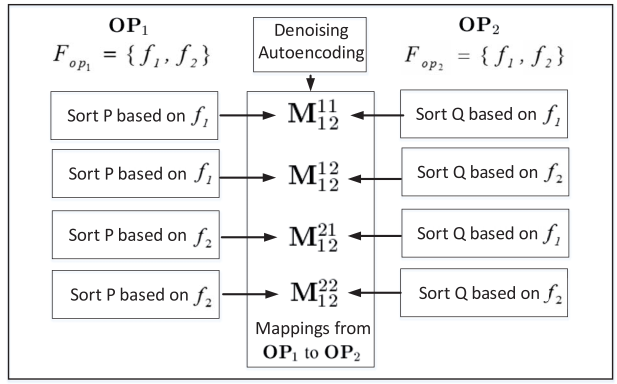


Fig. 3. Illustration of learning multiple mappings from $\mathbf{OP}_1$ to $\mathbf{OP}_2$ in MOO.

Algorithm 2: Pseudo Code of the Explicit Genetic Transfer Across Multiobjective Problem $\mathbf{OP}_1$ and $\mathbf{OP}_2$

Begin
Randomly select two objective functions that one is from $\mathbf{OP}_1$ and the other is from $\mathbf{OP}_2$, denoted by function a and function b, respectively.
Select fittest S solutions with respect to function a in the population of the evolutionary solver for $\mathbf{OP}_1$, and select another fittest S solutions with respect to function b in the population of the evolutionary solver for $\mathbf{OP}_2$.
Transfer the selected $\mathbf{OP}_1$ solutions to $\mathbf{OP}_2$ using the task mapping $\mathbf{M}_{12}^{ab}$, and transfer the selected $\mathbf{OP}_2$ solutions to $\mathbf{OP}_1$ using the task mapping $\mathbf{M}_{21}^{ba}$.
End

TABLE I
PROPERTIES OF SINGLE-OBJECTIVE MULTITASK OPTIMIZATION PROBLEMS

Category	Task	Search Space	Landscape	Degree of intersection	Inter-task similarity
CI+HS	Griewank (T1)	$D = 50, x \in [-100, 100]^D$	multimodal, nonseparable	Complete intersection	1.0000
	Rastrigin (T2)	$D = 50, x \in [-50, 50]^D$	multimodal, nonseparable		
CI+MS	Ackley (T1)	$D = 50, x \in [-50, 50]^D$	multimodal, nonseparable	Complete intersection	0.2261
	Rastrigin (T2)	$D = 50, x \in [-50, 50]^D$	multimodal, nonseparable		
CI+LS	Ackley (T1)	$D = 50, x \in [-50, 50]^D$	multimodal, nonseparable	Complete intersection	0.0002
	Schwefel (T2)	$D = 50, x \in [-500, 500]^D$	multimodal, separable		
PI+HS	Rastrigin (T1)	$D = 50, x \in [-500, 500]^D$	multimodal, nonseparable	Partial intersection	0.8670
	Sphere (T2)	$D = 50, x \in [-50, 50]^D$	unimodal, separable		
PI+MS	Ackley (T1)	$D = 50, x \in [-50, 50]^D$	multimodal, nonseparable	Partial intersection	0.2154
	Rosenbrock (T2)	$D = 50, x \in [-50, 50]^D$	multimodal, nonseparable		
PI+LS	Ackley (T1)	$D = 50, x \in [-50, 50]^D$	multimodal, nonseparable	Partial intersection	0.0725
	Weierstrass (T2)	$D = 25, x \in [-0.5, 0.5]^D$	multimodal, nonseparable		
NI+HS	Rosenbrock (T1)	$D = 50, x \in [-50, 50]^D$	multimodal, nonseparable	No intersection	0.9434
	Rastrigin (T2)	$D = 50, x \in [-50, 50]^D$	multimodal, nonseparable		
NI+MS	Griewank (T1)	$D = 50, x \in [-100, 100]^D$	multimodal, nonseparable	No intersection	0.3669
	Weierstrass (T2)	$D = 25, x \in [-0.5, 0.5]^D$	multimodal, nonseparable		
NI+LS	Rastrigin (T1)	$D = 50, x \in [-50, 50]^D$	multimodal, nonseparable	No intersection	0.0016
	Schwefel (T2)	$D = 50, x \in [-500, 500]^D$	multimodal, separable		

TABLE II
AVERAGED OBJECTIVE VALUE AND STANDARD DEVIATION OBTAINED BY THE PROPOSED EMT ALGORITHM, MFEA, GA, AND DE ON THE SINGLE-OBJECTIVE MULTITASK PROBLEM. ("$\approx$," "+," AND "$-$" DENOTE THE PROPOSED ALGORITHM STATISTICALLY SIGNIFICANT SIMILAR, BETTER, AND WORSE THAN MFEA, RESPECTIVELY)

Problem	Task No.	Proposed Method	MFEA	GA	DE
CI+HS	T1	**1.20E-5** ± 4.70E-5 +	3.65E-1 ± 6.73E-2	1.66E-1 ± 4.37E-2	–
	T2	**2.16E-2** ± 8.54E-2 +	1.81E+2 ± 4.67E+1	–	1.89E+2 ± 5.40E+1
CI+MS	T1	**2.77E-1** ± 5.18E-1 +	4.47E+0 ± 7.16E-1	3.12E+0 ± 2.75E-1	–
	T2	**1.37E+1** ± 2.46E+1 +	2.19E+2 ± 5.88E+1	–	1.78E+2 ± 5.16E+1
CI+LS	T1	2.10E+1 ± 4.14E-1 –	**2.02E+1** ± 5.28E-2	2.12E+1 ± 3.95E-2	–
	T2	6.73E+3 ± 9.17E+2 –	**3.69E+3** ± 6.11E+2	–	6.85E+3 ± 8.35E+2
PI+HS	T1	**3.15E+1** ± 5.54E+0 +	5.52E+2 ± 9.41E+1	4.31E+2 ± 6.32E+1	–
	T2	1.74E+2 ± 1.99E+2 –	**8.37E+0** ± 1.94E+0	–	3.06E+2 ± 2.16E+2
PI+MS	T1	**1.77E+0** ± 4.84E-1 +	3.65E+0 ± 8.93E-1	4.19E+0 ± 2.98E+0	–
	T2	**1.57E+2** ± 5.40E+1 +	6.80E+2 ± 3.44E+2	–	3.36E+5 ± 3.56E+5
PI+LS	T1	**2.30E-5** ± 4.80E-5 +	2.00E+1 ± 9.31E-2	3.51E+0 ± 5.84E-1	–
	T2	**2.35E-3** ± 3.93E-3 +	2.07E+1 ± 2.93E+0	–	2.27E+0 ± 1.07E+0
NI+HS	T1	**6.00E+1** ± 2.96E+1 +	9.88E+2 ± 6.85E+2	1.70E+3 ± 1.35E+3	–
	T2	**2.84E+1** ± 2.25E+1 +	2.56E+2 ± 6.67E+1	–	1.91E+2 ± 4.70E+1
NI+MS	T1	8.07E-1 ± 7.13E-2 –	**4.31E-1** ± 6.18E-2	8.65E-1 ± 8.12E-2	–
	T2	**2.63E+0** ± 5.15E-1 +	2.67E+1 ± 3.45E+0	–	1.28E+1 ± 1.87E+0
NI+LS	T1	**3.43E+1** ± 5.84E+0 +	6.14E+2 ± 1.17E+2	4.18E+2 ± 8.92E+1	–
	T2	7.08E+3 ± 9.85E+2 –	**3.69E+3** ± 4.56E+2	–	6.59E+3 ± 7.49E+2

IV. EMPIRICAL STUDY

In this section, comprehensive empirical studies have been conducted to evaluate the performance of the proposed EMT paradigm with *explicit genetic transfer*, on both single- and multi-objective multitask optimization problems.

A. Single-Objective Multitask Optimization

1) Experiment Setup: Based on the evolutionary multitask optimization technical report [17], nine single-objective multitask benchmark problems are employed in this paper. A summary of the problem properties is provided in Table I. As can be observed, this benchmark set comprises of test problems which are built by pairing classical single objective functions, such as *Griewank*, *Rastrigin*, *Schwefel*, etc., with the consideration of the function landscape type, degree of intersection between tasks, and intertask similarity.[4] For more details of the single-objective multitask benchmark set, interested reader is refer to [17].

[4]The intertask similarity is defined based on the Pearson correlation [17]. A higher value denotes greater similarity between tasks.

Next, the EMT algorithm MFEA proposed in [14] with *implicit genetic transfer* is considered as the baseline for comparison. According to [17], in MFEA, the genetic solver with simulated binary crossover (SBX) [40] and polynomial mutation are employed for offspring creation. Further, as there are two tasks in the single objective multitask benchmark, two simple basic single-task solvers, i.e., genetic algorithm (GA) and differential evolution (DE), are employed in the proposed EMT algorithm for the first and second task, respectively. For fair comparison, the SBX crossover and polynomial mutation are also configured in GA, while the common used *DE/rand/1* operator is used in DE. Other experimental settings are kept consistent with [17], and summarized as follows.

1) Number of sampled solutions for learning the task mappings: NoS $= 100$.
2) *Population Size:* Population size are configured as 100 for all the solvers.
3) *Maximum Generations:* $\text{Max}_{\text{Gen}} = 1000$ in MFEA, and $\text{Max}_{\text{Gen}} = 500$ for the others.
4) Independent number of runs: runs $= 20$.

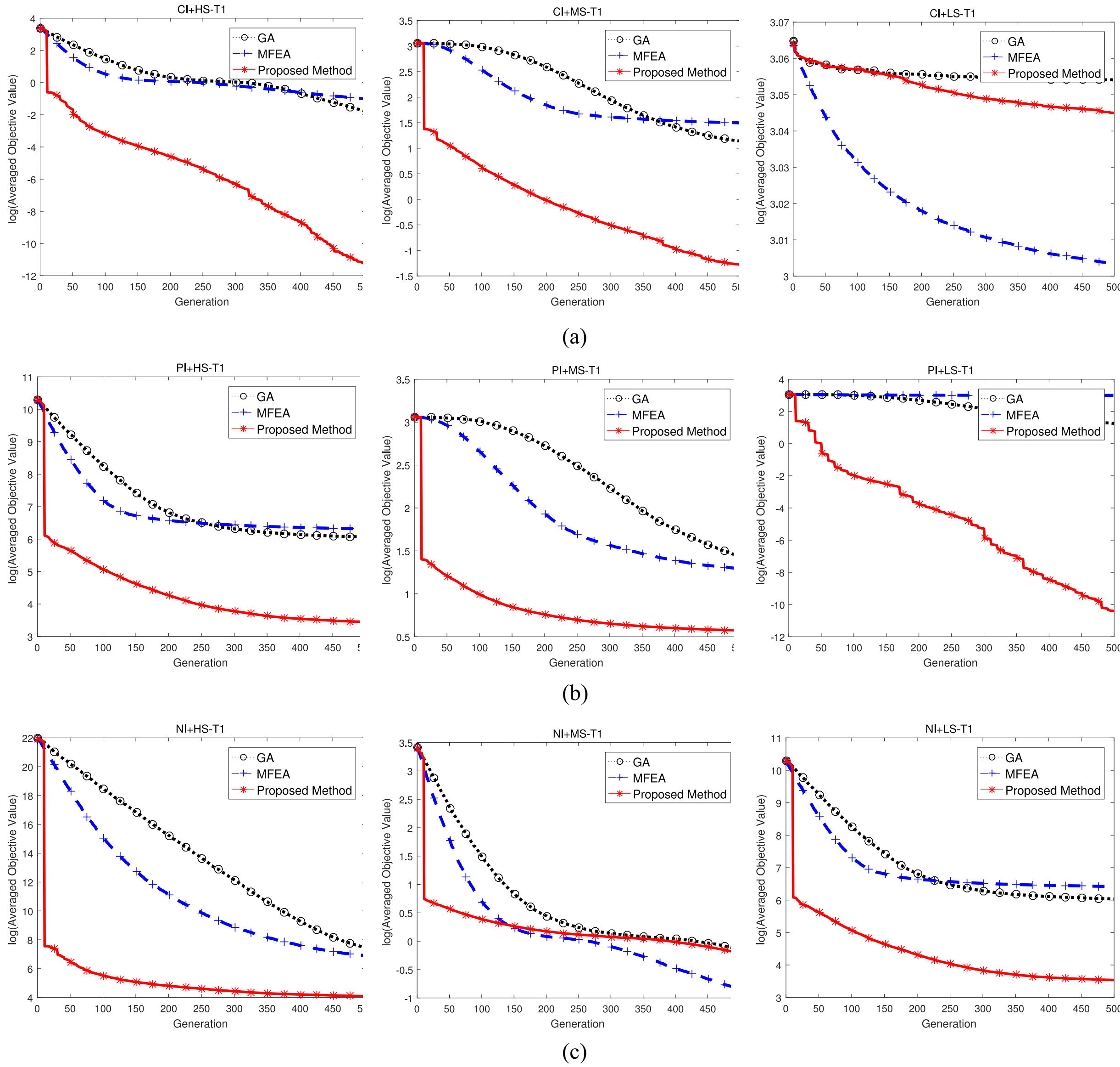


Fig. 4. Convergence traces of the proposed algorithm, MFEA, and single-task GA on T1 of the nine single-objective multitask problems. y-axis: log(objective value); x-axis: generation. Convergence traces of T1 in (a) CI+HS, CI+MS, and CI+LS, (b) PI+HS, PI+MS, and PI+LS, and (c) NI+HS, NI+MS, and NI+LS.

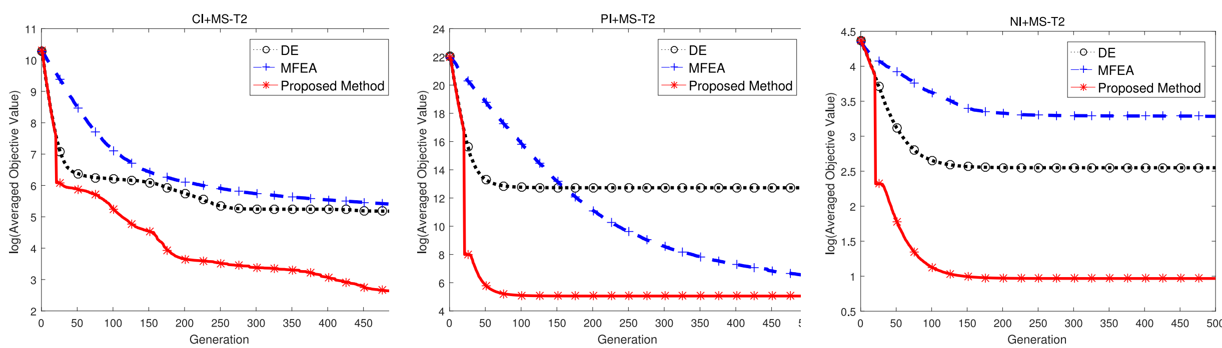


Fig. 5. Convergence traces of the proposed algorithm, MFEA, and single-task DE on T2 of benchmark CI+MS, PI+MS, and NI+MS. y-axis: log(objective value); x-axis: generation.

5) Evolutionary operators and parameters in the proposed algorithm, MFEA, GA, and DE.
   a) *SBX Crossover:* $p_c = 1$, $\eta_c = 2$.
   b) *Polynomial Mutation:* $p_m = 1$, $\eta_m = 5$.
6) *rmp* in MFEA: $rmp = 0.3$.
7) F and CR in DE: $F = 0.5$, CR $= 0.6$.

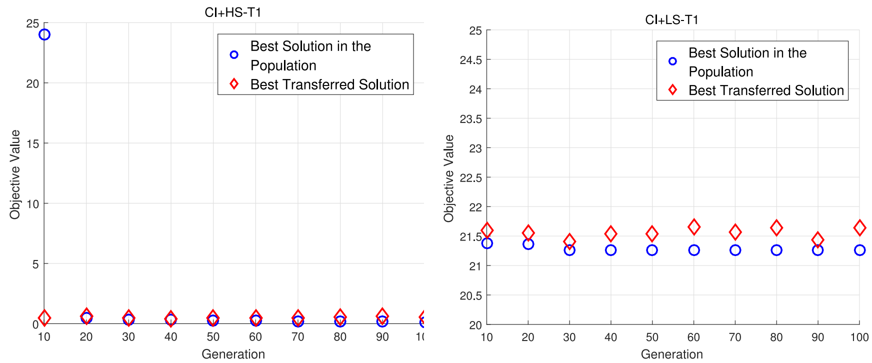


Fig. 6. Illustration of the transferred solutions and the best solution in the population on representative single-objective benchmarks.

8) Interval of explicit solution transfer across tasks: $G = 10$.
9) Solutions to be transferred across tasks: $S = 10$.

As can be observed, since MFEA has only one population for solving two tasks, the maximal number of generation in MFEA is doubled when compared to the other algorithms. The parameter setting of DE is referred to [41]. Further, the explicit solution transfer across tasks in the proposed EMT algorithm happens in every ten generation, and ten best solutions in terms of objective value from each tasks will be selection for transfer from one to another. Lastly, to be consistent with the setting in [17], no separate local search steps are performed in this paper.

2) Results and Discussions: Table II summarizes the averaged objective value and standard deviation obtained by the proposed EMT algorithm, MFEA, GA, and DE over 20 independent runs on the single-objective multitask optimization benchmarks. Since GA and DE have been employed as the two basic solvers in the proposed algorithm for tackling T1 and T2 in each benchmark, respectively, the results obtained by single-task GA on T1 and single-task DE on T2 of each problem are also presented in the table for comparison. Further, in the table, superior performance is highlighted in bold, and the Wilcoxon rank sum test with 95% confidence level is conducted on the experimental results. $\approx$, $+$, and $-$ denote the proposed algorithm statistically significant similar, better, and worse than MFEA, respectively.

As can be observed in the table, first of all, with explicit genetic solution transfer across tasks, compared to the single-task GA and DE, the proposed EMT achieved superior performance on all the benchmarks in terms of averaged objective value. As the proposed EMT algorithm share the same evolutionary solver with GA and DE, and the only difference lies in the explicit genetic solution transfer across tasks in the proposed method, the obtained improved solution quality confirmed the effectiveness of conducting EMT for optimization. Further, compared to the MFEA, the proposed algorithm obtained significantly superior solution quality in terms of averaged objective value on 13 out of 18 tasks, which further confirmed the effectiveness of the proposed *explicit genetic transfer* across task for EMT. In particular, the proposed algorithm achieved solutions approaching to the global optimum 0 on T2 of CI+HS and the two tasks of PI+LS, where MFEA converged to local optima.

TABLE III
PROPERTIES OF MULTIOBJECTIVE MULTITASK OPTIMIZATION PROBLEMS

Problem	*sim* (T1, T2)	Task No.	Pareto Set	Pareto Front	Properties
CIHS	0.97	T1	$x_1 \in [0,1]$, $x_i = 0, i = 2:50$	$f_1^2 + f_2^2 = 1$, $f_1 \geq 0, f_2 \geq 0$	concave, unimodal, separable
		T2	$x_1 \in [0,1]$, $x_i = 0, i = 2:50$	$f_2 = 1 - f_1^2$, $0 \leq f_1 \leq 1$	concave, unimodal, separable
CIMS	0.52	T1	$x_1 \in [0,1]$, $x_i = 1, i = 2:10$	$f_2 = 1 - f_1^2$, $0 \leq f_1 \leq 1$	concave, multimodal, nonseparable
		T2	$x_1 \in [0,1]$, $(x_2, \ldots, x_{10})^{\mathrm{T}} = \mathbf{s}_{cm2}$	$f_1^2 + f_2^2 = 1$, $f_1 \geq 0, f_2 \geq 0$	concave, unimodal, nonseparable
CILS	0.07	T1	$x_1 \in [0,1]$, $x_i = 0, i = 2:50$	$f_1^2 + f_2^2 = 1$, $f_1 \geq 0, f_2 \geq 0$	concave, multimodal, separable
		T2	$x_1 \in [0,1]$, $x_i = 0, i = 2:50$	$f_2 = 1 - \sqrt{f_1}$, $0 \leq f_1 \leq 1$	convex, multimodal, nonseparable
PIHS	0.99	T1	$x_1 \in [0,1]$, $x_i = 0, i = 2:50$	$f_2 = 1 - \sqrt{f_1}$, $0 \leq f_1 \leq 1$	convex, unimodal, separable
		T2	$x_1 \in [0,1]$, $(x_2, \ldots, x_{50})^{\mathrm{T}} = \mathbf{s}_{ph2}$	$f_2 = 1 - \sqrt{f_1}$, $0 \leq f_1 \leq 1$	convex, multimodal, separable
PIMS	0.55	T1	$x_1 \in [0,1]$, $(x_2, \ldots, x_{50})^{\mathrm{T}} = \mathbf{s}_{pm1}$	$f_1^2 + f_2^2 = 1$, $f_1 \geq 0, f_2 \geq 0$	concave, unimodal, nonseparable
		T2	$x_1 \in [0,1]$, $x_i = 0, i = 2:50$	$f_2 = 1 - f_1^2$, $0 \leq f_1 \leq 1$	concave, multimodal, nonseparable
PILS	0.002	T1	$x_1 \in [0,1]$, $x_i = 0, i = 2:50$	$f_1^2 + f_2^2 = 1$, $f_1 \geq 0, f_2 \geq 0$	concave, multimodal, nonseparable
		T2	$x_1 \in [0,1]$, $(x_2, \ldots, x_{50})^{\mathrm{T}} = \mathbf{s}_{pl2}$	$f_1^2 + f_2^2 = 1$, $f_1 \geq 0, f_2 \geq 0$	concave, multimodal, nonseparable
NIHS	0.94	T1	$x_1 \in [0,1]$, $x_i = 1, i = 2:50$	$f_1^2 + f_2^2 = 1$, $f_1 \geq 0, f_2 \geq 0$	concave, multimodal, nonseparable
		T2	$x_1 \in [0,1]$, $x_i = 0, i = 2:50$	$f_2 = 1 - \sqrt{f_1}$, $0 \leq f_1 \leq 1$	convex, unimodal, separable
NIMS	0.51	T1	$x_1 \in [0,1], x_2 \in [0,1]$, $x_i = 1, i = 3:20$	$\sum_{i=1}^{3} f_i^2 = 1$, $f_i \geq 0, i = 1,2,3$	concave, multimodal, nonseparable
		T2	$x_1 \in [0,1], x_2 \in [0,1]$, $x_i = 0, i = 3:20$	$f_2 = 1 - f_1^2$, $0 \leq f_1 \leq 1$	concave, unimodal, nonseparable
NILS	0.001	T1	$x_1 \in [0,1], x_2 \in [0,1]$, $(x_3, \ldots, x_{25})^{\mathrm{T}} = \mathbf{s}_{nl1}$	$\sum_{i=1}^{3} f_i^2 = 1$, $f_i \geq 0, i = 1,2,3$	concave, multimodal, nonseparable
		T2	$x_1 \in [0,1], x_2 \in [0,1]$, $x_i = 0, i = 3:50$	$f_2 = 1 - f_1^2$, $0 \leq f_1 \leq 1$	concave, multimodal, nonseparable

Next, to access the efficiency of our proposed approach, the average convergence graphs on the single-objective multitask benchmarks are depicted in Fig. 4. In particular, as MFEA shares the same underlying evolutionary operators with GA, to clearly validate the efficiency of both the EMT and the proposed explicit genetic transfer across tasks, the convergence trances obtained by the proposed algorithm, MFEA and GA, on task T1 of all the nine single-objective multitask problems

TABLE IV
AVERAGED VALUE AND STANDARD DEVIATION OF THE IGD OBTAINED BY THE PROPOSED EMT ALGORITHM, MOMFEA, SPEA2, AND NSGAII, ON THE NINE MULTIOBJECTIVE MULTITASK BENCHMARKS. ($\approx$, $+$, AND $-$ DENOTE THE PROPOSED ALGORITHM STATISTICALLY SIGNIFICANT SIMILAR, BETTER, AND WORSE THAN MFEA, RESPECTIVELY

Problem	Task No.	Proposed Method	MOMFEA	SPEA2	NSGAII
CIHS	T1	**1.49E-4** ± 3.00E-6 +	4.15E-4 ± 1.18E-4	2.10E-3 ± 5.35E-4	–
	T2	**2.02E-4** ± 3.70E-5 +	2.74E-3 ± 4.05E-4	–	4.53E-3 ± 6.26E-4
CIMS	T1	1.70E-1 ± 4.94E-2 –	**3.89E-2** ± 6.91E-2	1.26E-1 ± 7.99E-2	–
	T2	2.01E-2 ± 1.51E-2 –	**6.59E-3** ± 8.30E-3	–	2.64E-2 ± 2.48E-2
CILS	T1	**1.49E-4** ± 3.00E-6 +	2.71E-4 ± 4.00E-5	2.69E-1 ± 9.28E-2	–
	T2	**1.86E-4** ± 9.00E-6 ≈	1.88E-4 ± 6.00E-6	–	2.01E-4 ± 7.00E-6
PIHS	T1	**9.80E-4** ± 3.22E-4 ≈	1.16E-3 ± 1.07E-3	1.34E-3 ± 3.40E-4	–
	T2	3.94E-2 ± 1.94E-2 ≈	**3.85E-2** ± 1.23E-2	–	5.73E-2 ± 3.30E-2
PIMS	T1	**1.75E-3** ± 1.18E-3 +	2.97E-3 ± 1.44E-3	5.68E-3 ± 1.87E-3	–
	T2	**2.44E+0** ± 1.85E+0 +	1.09E+1 ± 3.76E+0	–	1.58E+1 ± 2.82E+0
PILS	T1	**2.27E-4** ± 1.14E-4 +	3.88E-4 ± 1.70E-4	2.57E-4 ± 1.00E-4	–
	T2	6.26E-2 ± 1.51E-1 –	**1.11E-2** ± 2.62E-3	–	6.36E-1 ± 1.69E-3
NIHS	T1	**1.50E+0** ± 6.36E-3 +	1.55E+0 ± 1.61E-2	1.09E+1 ± 1.85E+1	–
	T2	**1.92E-4** ± 1.30E-5 +	4.76E-4 ± 1.13E-4	–	8.12E-4 ± 1.99E-4
NIMS	T1	**1.61E-1** ± 3.25E-3 +	2.03E-1 ± 2.12E-1	4.32E-1 ± 3.40E-1	–
	T2	**1.91E-4** ± 1.10E-5 +	4.04E-2 ± 5.78E-2	–	7.90E-2 ± 6.23E-2
NILS	T1	6.40E-4 ± 3.30E-5 +	8.61E-4 ± 5.90E-5	**6.33E-4** ± 2.50E-5	–
	T2	**3.89E-1** ± 3.15E-1 +	6.43E-1 ± 3.41E-4	–	6.42E-1 ± 2.14E-4

are presented. In the figure, y-axis denotes the averaged objective values in log scale, while the x-axis gives the respective computational effort incurred in terms of the number of generations. It can be observed from Fig. 4 that, with knowledge sharing across tasks, both MFEA and the proposed algorithm converge faster over the single-task GA on most of the problems. On the other hand, we observe that the proposed EMT algorithm converges much fast than MFEA on almost all the problems, except T1 of CI+LS which has very low similarity with T2 in CI+LS (see Table I). On tasks, such as T1 of CI+HS, CI+MS, PI+MS, NI+HS, etc., a clear drop of objective value can be observed in the convergence traces of the proposed method, which could be contributed to the explicit solution transfer across tasks.

We further depict the convergence traces obtained by the proposed method, MFEA, and DE on task T2 of CI+MS, PI+MS, and NI+MS in Fig. 5. As DE has different evolutionary operators from MFEA, it can be observed from Fig. 5 that, DE searches faster than MFEA. With the proposed method, the faster search performance of DE can be transferred to improve the search of GA on the corresponding task T1, thus lead to much enhanced convergence speed of the proposed algorithm as observed in Fig. 4. This again confirmed the efficacy of the proposed EMT algorithm with *explicit genetic transfer* across tasks. Further, in Fig. 5, with solutions transferred from the GA on task T1, the search performance of the proposed algorithm has also been improved on the respective task T2, when compared to the single-task DE.

Lastly, to provide deeper insight of the obtained performance of our proposed EMT paradigm. An illustration of the transferred solutions via the proposed single-layer denoising autoencoding and the best solutions in the population on representative single-objective benchmarks is presented in Fig. 6. As can be observed, the transferred solution for the task T1 of CI+HS in Fig. 6 explained the superior performance obtained by the proposed EMT on T1 of CI+HS in Fig. 4(a). In particular, a high-quality solution approaching to the global optimum is transferred to T1 in generation 10, and thus lead the search toward the global optimal solution in the subsequent generations. On the other hand, since the transferred solutions are close to the best solutions in the population on task T1 of CI+LS (see Fig. 6), the proposed EMT performs competitively against the single-task solver, as depicted in Fig. 4(a).

B. Multiobjective Multitask Optimization

1) Experiment Setup: Similar to the study conducted on single-objective multitask optimization above, nine multiobjective multitask problems built in the recent technical report [18] are considered in this experimental study. In particular, the multiobjective multitask problems are built by considering the degree of the intersection of the Pareto optimal solutions and the similarity in the fitness landscape between tasks. Each problem contains two MOO problems with 2 or 3 objectives that are commonly studied in the literature. The detailed problem properties, including, task similarity (i.e., sim),[5] Pareto set, Pareto front, etc., are summarized in Table III. For more introduction of developing the multiobjective multitask benchmarks, interested readers may refer to [18].

Subsequently, the recently proposed MOMFEA [19] with *implicit genetic transfer* for multiobjective multitask optimization is considered here as the baseline algorithm. In the proposed EMT algorithm, the well-known NSGAII [42] is employed as one of the multiobjective solver, since it has also been used as the underlying basic evolutionary search in MOMFEA. Further, the SPEA2 [43] is considered as the other basic solver in our proposed method. Particular parameter settings are configured based on [19], which are summarized below.

1) Number of sampled solutions for learning the task mappings: NoS $= 100$.

[5]The intertask similarity is defined based on the Pearson correlation [18]. A higher value denotes greater similarity between tasks.

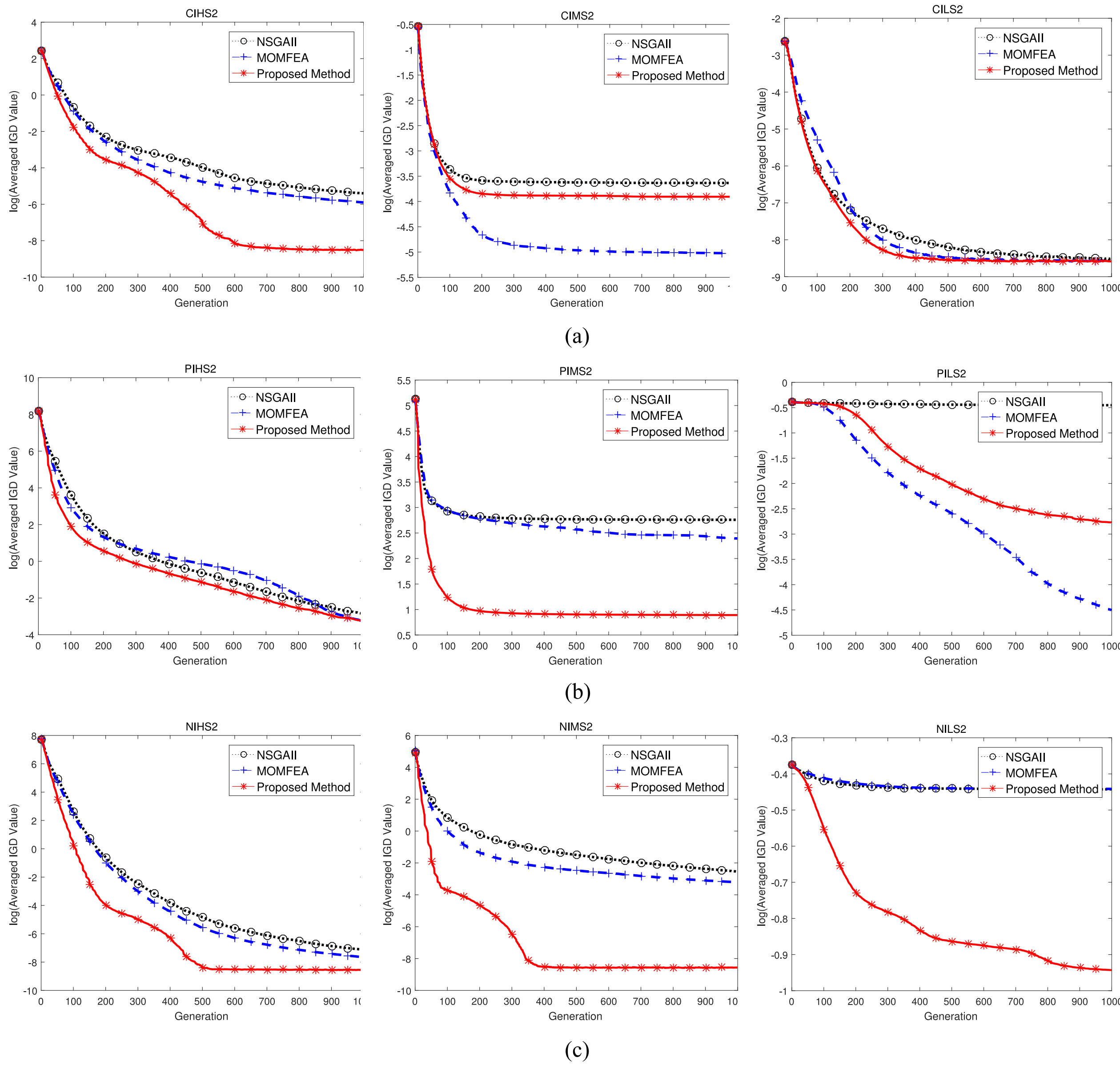


Fig. 7. Convergence curves of the averaged IGD (over 30 runs) obtained by the proposed EMT algorithm, MOMFEA, and NSGAII on T2 of the nine multiobjective multitask problems. *y*-axis: log(IGD value); *x*-axis: generation. Convergence curves of averaged IGD on T2 in (a) CIHS, CIMS, and CILS, (b) PIHS, PIMS, and PILS, and (c) NIHS, NIMS, and NILS.

2) *Population Size:* Population size of each solver, i.e., NSGAII and SPEA2, in the proposed algorithm is configured as 100, while the population size of MOMFEA is set as 200.
3) *Maximum Generations:* $\text{Max}_{\text{Gen}} = 1000$.
4) Independent number of runs: runs $= 30$.
5) Evolutionary operators and parameters in MOMFEA, the proposed EMT algorithm, NSGAII, and SPEA2.
   a) *SBX Crossover:* $p_c = 0.9$, $\eta_c = 20$.
   b) *Polynomial Mutation:* $p_m = 1/D$,[6] $\eta_m = 20$.
6) Interval of explicit solution transfer across tasks: $G = 10$.
7) Solutions to be transferred across tasks: $S = 10$.

[6]*D* is the dimensionality of the encoded solution. For MOMFEA, it is the dimensionality of the unified code representation

As MOMFEA has only one population for solving two tasks, for fair comparison, its population is set as double of that for the proposed EMT algorithm. Further, the frequency and the number of solutions for explicit transfer across tasks is kept the same as that in the single-objective multitask optimization study. However, the selection of the ten best solution in the context of MOO is executed as discussed in Section III-B2.

2) Results and Discussions: To evaluate the performance of the proposed EMT algorithm on multiobjective multitask optimization problems, the averaged IGD values and the standard deviations obtained by the proposed algorithm, MOMFEA, SPEA2, and NSGAII, over 30 independent runs on the multiobjective multitask benchmarks, are summarized in Table IV. In the table, superior performance is highlighted in bold, and the Wilcoxon rank sum test with 95% confidence level is also conducted on the experimental results. $\approx$, $+$, and

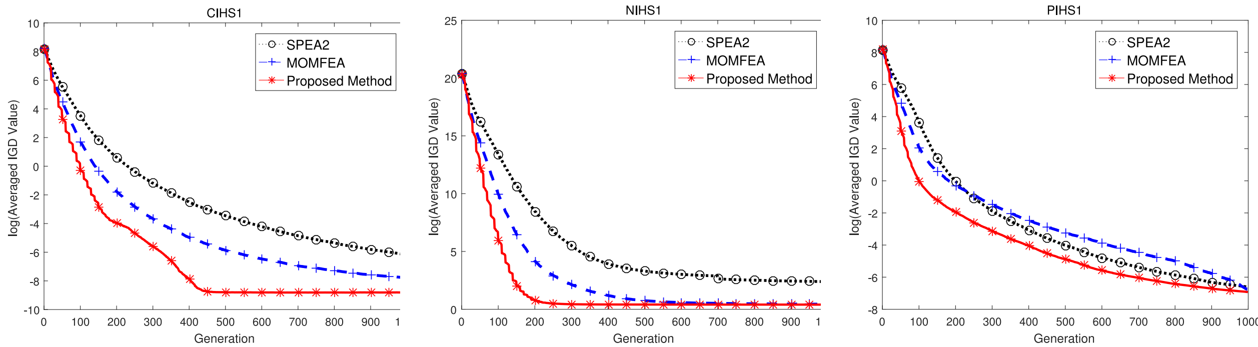


Fig. 8. Convergence curves of the averaged IGD (over 30 runs) obtained by the proposed EMT algorithm, MOMFEA, and SPEA2 on T1 of representative multiobjective multitask problems. y-axis: log(IGD value); x-axis: generation.

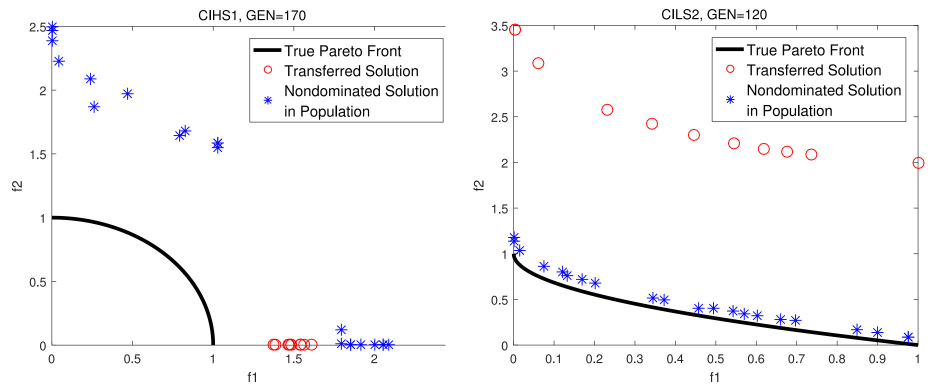


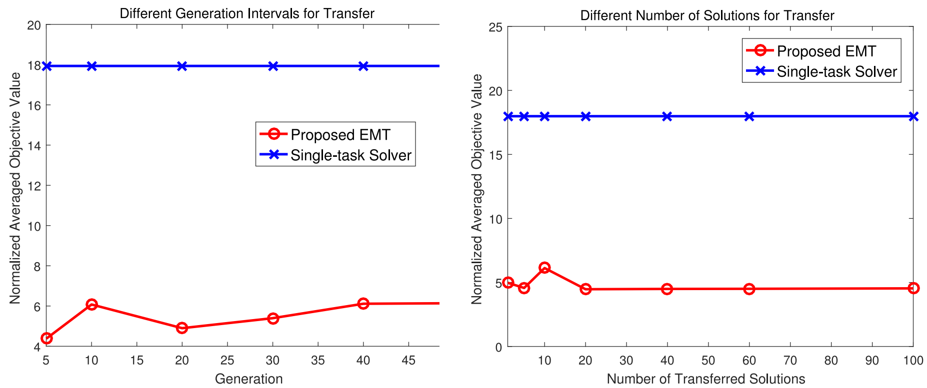
Fig. 9. Illustration of the transferred solutions and the best solution in the population on representative multiobjective benchmarks.

$-$ denote the proposed algorithm statistically significant similar, better, and worse than MFEA, respectively. Further, please note that, in both the proposed algorithm and the single-task evolutionary search, SPEA2 and NSGAII are used for solving T1 and T2 in each multitask benchmark, respectively.

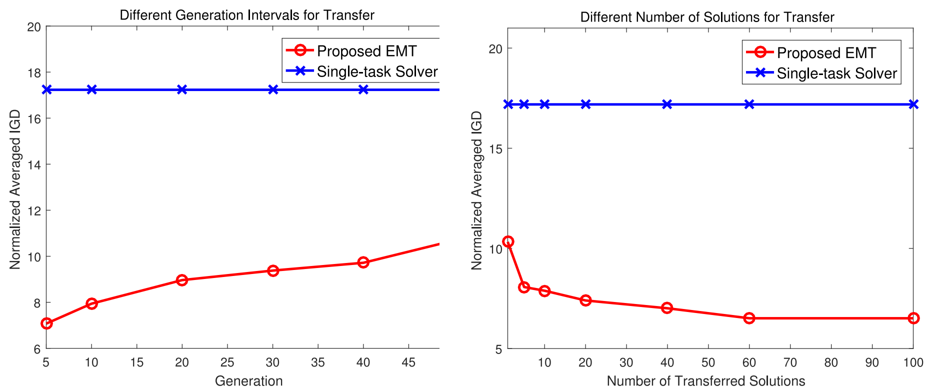
As can be observed, with knowledge sharing between tasks, the two EMT algorithms (i.e., the proposed method and MOMFEA) achieved superior performance over the single task SPEA2 and NSGAII on almost all the multitasking benchmarks, in terms of IGD value. In particular, on task T1 of CILS, the single-task SPEA2 obtained the IGD of $2.69E-1$, while the proposed method and MOMFEA achieved $1.49E-4$ and $2.71E-4$, respectively. On task T2 of PILS, the proposed method and MOMFEA obtained the IGD value of $6.26E-2$ and $1.11E-2$, respectively, while the single-task NSGAII only researched the IGD $6.36E-1$. Further, with respect to the EMT algorithms, the proposed algorithm obtained competitive and significant superior IGD value against the MOMFEA on 15 of totally 18 multiobjective tasks.

Subsequently, Fig. 7 provides the averaged convergence graphs obtained by the proposed algorithm, the MOMFEA, and NSGAII, on T2 of all the multiobjective multitask benchmarks. In the figure, y-axis gives the averaged IGD value in log scale, while the x-axis denotes the number of generations made so far. It can be observed from the figure that, even MOMFEA and NSGAII share the same evolutionary search operators, with the implicit genetic transfer across tasks, MOMFEA demonstrated faster or competitive convergence speed over NSGAII on the multiobjective tasks. For our proposed algorithm, it converges faster than the MOMFEA on most of the tasks, e.g., CIHS2 in Fig. 7(a), PIMS2 in Fig. 7(b), NILS2 in Fig. 7(c), etc. This is because the NSGAII in the proposed algorithm for T2 has solutions transferred from the SPEA2 for T1 with the proposed explicit solution transfer between tasks. On the CIMS2 [in Fig. 7(a)] and PILS2 [in Fig. 7(b)], due to the relatively low similarity between tasks, deteriorated convergence speed of the proposed method has also been observed against MOMFEA. Further, similar observations have also been achieved by the proposed method when compared to the MOMFEA on T1 of the multiobjective multitask benchmarks. Representative convergence curves of the averaged IGD obtained by the proposed EMT algorithm, MOMFEA and SPEA2 are presented in Fig. 8.

Further, the illustration of the transferred solutions via the proposed single-layer denoising autoencoder and the best solutions in the population on representative multiobjective benchmarks is presented in Fig. 9, to give deeper insights of the obtained performance of our proposed EMT paradigm. As can be observed, in Fig. 9, on task T2 of CI+HS benchmark, as solutions approaching to the true pareto front have been transferred across tasks based on the randomly selected objective (see Section III-B2), enhanced search performance of the proposed EMT can be observed in Fig. 7(a). However, since the transferred solution for task T2 of CI+LS are dominated by the existing solutions in the population (see Fig. 9), we observed



(a)



(b)

Fig. 10. Averaged objective value and averaged IGD obtained by the proposed EMT and the single-task solvers on all the single- and multi-objective benchmarks across 20 independent runs with various configuration of G and S.

that the proposed EMT performs competitively against the single-task solver in Fig. 7(a).

Lastly, as G and S define the frequency and amount of knowledge sharing between tasks, we further study how the configurations of G and S affect the proposed EMT paradigm. Generally, a small value of G and a big value of S will greatly increase the frequency and amount of knowledge sharing across tasks, while a big value of G and a small value of S will reduce the frequency and amount of solution transfer across tasks significantly. Fig. 10 gives the summation of normalized average objective value and average IGD obtained by the proposed method on all the single- and multi-objective benchmarks across 20 independent runs with various configuration of G and S. It can be observed from the figure, superior solution qualities have been obtained by the proposed EMT paradigm when compared to the single-task solvers on both single- and multi-objective problems with all the configurations of G and S. However, while the optimal confirmations of G and S are in general problem dependent, fixing $G = 10$ and $S = 10$ is found to provide noteworthy results across a variety of problems encountered.

V. Conclusion

This paper has proposed a new mode of knowledge transfer across tasks in EMT via a denoising autoencoder for both single- and multi-objective multitask optimization. In contrast to existing EMT algorithms which conduct knowledge transfer *implicitly* through the genetic crossover, the proposed method performs solution transfer among tasks *explicitly* using a single layer denoising autoencoder with a closed-form solution. It allows the incorporation of different search mechanisms with unique bias for EMT. To validate the performance of the proposed EMT algorithm, comprehensive empirical studies on both the single- and multi-objective multitasking benchmarks have been conducted. The results obtained have been compared with recently proposed EMT algorithms and the corresponding single-task evolutionary solvers, which confirmed the efficacy of the proposed EMT algorithm with explicit genetic transfer.

For future works, we would like to further study the proposed EMT via explicit autoencoding on more complex and high-dimensional optimization problems. The design of other efficient learning methods for conducting explicit knowledge transfer across tasks along the evolutionary search process is also a promising research direction.

References

[1] T. Bäck, U. Hammel, and H.-P. Schwefel, "Evolutionary computation: Comments on the history and current state," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 3–17, Apr. 1997.

[2] H. Maaranen, K. Miettinen, and A. Penttinen, "On initial populations of a genetic algorithm for continuous optimization problems," *J. Glob. Optim.*, vol. 37, no. 3, pp. 405–436, 2007.

[3] D. K. Saxena, A. Sinha, J. A. Duro, and Q. Zhang, "Entropy-based termination criterion for multiobjective evolutionary algorithms," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 485–498, Aug. 2016.

[4] S. Jiang and S. Yang, "Evolutionary dynamic multiobjective optimization: Benchmarks and algorithm comparisons," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 198–211, Jan. 2017.

[5] X. Ma, Q. Zhang, G. Tian, J. Yang, and Z. Zhu, "On Tchebycheff decomposition approaches for multiobjective evolutionary optimization," *IEEE Trans. Evol. Comput.*, vol. 22, no. 2, pp. 226–244, Apr. 2018.

[6] K. C. Tan, Y. H. Chew, and L. Lee, "A hybrid multiobjective evolutionary algorithm for solving vehicle routing problems with time windows," *Comput. Optim. Appl.*, vol. 34, no. 1, pp. 115–151, 2006.

[7] Z. Zhu, J. Xiao, S. He, Z. Ji, and Y. Sun, "A multi-objective memetic algorithm based on locality-sensitive hashing for one-to-many-to-one dynamic pickup-and-delivery problem," *Inf. Sci.*, vol. 329, pp. 73–89, Feb. 2016.

[8] M. Gong, H. Li, E. Luo, J. Liu, and J. Liu, "A multiobjective cooperative coevolutionary algorithm for hyperspectral sparse unmixing," *IEEE Trans. Evol. Comput.*, vol. 21, no. 2, pp. 234–248, Apr. 2017.

[9] K. Tang, Y. Mei, and X. Yao, "Memetic algorithm with extended neighborhood search for capacitated arc routing problems," *IEEE Trans. Evol. Comput.*, vol. 13, no. 5, pp. 1159–1166, Oct. 2009.

[10] F. Neri, C. Cotta, and P. Moscato, *Handbook of Memetic Algorithms* (Studies in Computational Intelligence). Heidelberg, Germany: Springer, 2011.

[11] Y.-L. Li *et al.*, "Differential evolution with an evolution path: A DEEP evolutionary algorithm," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1798–1810, Sep. 2015.

[12] Z. Zhu, F. Wang, S. He, and Y. Sun, "Global path planning of mobile robots using a memetic algorithm," *Int. J. Syst. Sci.*, vol. 46, no. 11, pp. 1982–1993, 2015.

[13] L. Cui *et al.*, "A novel differential evolution algorithm with a self-adaptation parameter control method by differential evolution," *Soft Comput.*, pp. 1–20, Jun. 2017, doi: 10.1007/s00500-017-2685-5.

[14] A. Gupta, Y.-S. Ong, and L. Feng, "Multifactorial evolution: Toward evolutionary multitasking," *IEEE Trans. Evol. Comput.*, vol. 20, no. 3, pp. 343–357, Jun. 2016.

[15] Y.-S. Ong and A. Gupta, "Evolutionary multitasking: A computer science view of cognitive multitasking," *Cogn. Comput.*, vol. 8, no. 2, pp. 125–142, 2016.

[16] A. Gupta, Y.-S. Ong, B. Da, L. Feng, and S. D. Handoko, "Landscape synergy in evolutionary multitasking," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Vancouver, BC, Canada, 2016, pp. 3076–3083.

[17] B. S. Da *et al.*, "Evolutionary multitasking for single-objective continuous optimization: Benchmark problems, performance metrics and baseline results," School Comput. Sci. Eng., Nanyang Technol. Univ., Singapore, Rep., 2016.

[18] Y. Yuan *et al.*, "Evolutionary multitasking for multiobjective continuous optimization: Benchmark problems, performance metrics and baseline results," School Comput. Sci. Eng., Nanyang Technol. Univ., Singapore, Rep., 2016.

[19] A. Gupta, Y.-S. Ong, L. Feng, and K. C. Tan, "Multi-objective multifactorial optimization in evolutionary multitasking," *IEEE Trans. Cybern.*, vol. 47, no. 7, pp. 1652–1665, Jul. 2017.

[20] L. Zhou *et al.*, "Evolutionary multitasking in combinatorial search spaces: A case study in capacitated vehicle routing problem," in *Proc. IEEE Symp. Series Comput. Intell. (SSCI)*, Athens, Greece, 2016, pp. 1–8.

[21] Y.-W. Wen and C.-K. Ting, "Parting ways and reallocating resources in evolutionary multitasking," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, 2017, pp. 2404–2411.

[22] R.-T. Liaw and C.-K. Ting, "Evolutionary many-tasking based on biocoenosis through symbiosis: A framework and benchmark problems," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, 2017, pp. 2266–2273.

[23] K. K. Bali, A. Gupta, L. Feng, Y.-S. Ong, and T. P. Siew, "Linearized domain adaptation in evolutionary multitasking," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, 2017, pp. 1295–1302.

[24] Z. Tang, M. Gong, and M. Zhang, "Evolutionary multi-task learning for modular extremal learning machine," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, 2017, pp. 474–479.

[25] M. Iqbal, B. Xue, H. Al-Sahaf, and M. Zhang, "Cross-domain reuse of extracted knowledge in genetic programming for image classification," *IEEE Trans. Evol. Comput.*, vol. 21, no. 4, pp. 569–587, Aug. 2017.

[26] D. O'Neill, H. Al-Sahaf, B. Xue, and M. Zhang, "Common subtrees in related problems: A novel transfer learning approach for genetic programming," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, 2017, pp. 1287–1294.

[27] L. J. Eshelman, R. Caruana, and J. D. Schaffer, "Biases in the crossover landscape," in *Proc. 3rd Int. Conf. Genet. Algorithms*, 1989, pp. 10–19.

[28] K. F. Man, K. S. Tang, and S. Kwong, "Genetic algorithms: Concepts and applications [in engineering design]," *IEEE Trans. Ind. Electron.*, vol. 43, no. 5, pp. 519–534, Oct. 1996.

[29] X. Zhang, Y. Zhuang, W. Wang, and W. Pedrycz, "Transfer boosting with synthetic instances for class imbalanced object recognition," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 357–370, Jan. 2018.

[30] L. Feng, Y.-S. Ong, M.-H. Lim, and I. W. Tsang, "Memetic search with interdomain learning: A realization between CVRP and CARP," *IEEE Trans. Evol. Comput.*, vol. 19, no. 5, pp. 644–658, Oct. 2015.

[31] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, U.K.: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[33] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, "Joint learning of words and meaning representations for open-text semantic parsing," in *Proc. Int. Conf. Artif. Intell. Stat.*, 2012, pp. 127–135.

[34] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proc. 28th Int. Conf. Mach. Learn. (ICML)*, Bellevue, WA, USA, 2011, pp. 97–110.

[35] L. Feng, Y.-S. Ong, S. Jiang, and A. Gupta, "Autoencoding evolutionary search with learning across heterogeneous problems," *IEEE Trans. Evol. Comput.*, vol. 21, no. 5, pp. 760–772, Oct. 2017.

[36] A. Graves *et al.*, "A novel connectionist system for unconstrained handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 855–868, May 2009.

[37] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, Mar. 2010.

[38] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[39] M. A. Potter and K. A. D. Jong, "Cooperative coevolution: An architecture for evolving coadapted subcomponents," *Evol. Comput.*, vol. 8, no. 1, pp. 1–29, 2000.

[40] K. Deb and R. B. Agrawal, "Simulated binary crossover for continuous search space," *Complex Syst.*, vol. 9, no. 2, pp. 1–15, 1994.

[41] R. Storn and K. Price, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[42] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[43] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," in *Proc. Evol. Methods Design Optim. Control Appl. Ind. Problems (EUROGEN)*, 2001, pp. 95–100.

[44] K. K. Bali, A. Gupta, L. Feng, Y. S. Ong, and P. S. Tan, "Linearized domain adaptation in evolutionary multitasking," in *Proc. IEEE Congr. Evol. Comput.*, 2017, pp. 1295–1302.

Liang Feng received the Ph.D. degree from the School of Computer Engineering, Nanyang Technological University, Singapore, in 2014.

He was a Post-Doctoral Research Fellow with the Computational Intelligence Graduate Laboratory, Nanyang Technological University. He is currently an Assistant Professor with the College of Computer Science, Chongqing University, Chongqing, China. His current research interests include computational and artificial intelligence, memetic computing, big data optimization and learning, as well as transfer learning.

Lei Zhou received the B.E. degree from the Department of Computer Science, Shandong University, Shandong, China, in 2014. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Chongqing University, Chongqing, China.

His current research interests include evolutionary computations, memetic computing, and transfer learning.

Jinghui Zhong received the Ph.D. degree from the School of Information Science and Technology, Sun Yat-sen University, Guangzhou, China, in 2012.

He is an Associate Professor with the School of Computer and Science Engineering, South China University of Technology, Guangzhou, China. His current research interests include computational intelligence, machine learning, and agent-based modeling and simulation.

Abhishek Gupta received the Bachelor of Technology degree from the National Institute of Technology, Rourkela, Rourkela, India, in 2010, and the Ph.D. degree in engineering science from the University of Auckland, Auckland, New Zealand, in 2014.

He currently serves as a Research Scientist with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He has diverse research experience in the field of computational science, ranging from the numerical modeling of solids and fluids, to topics in computational intelligence. His current research interests include development of memetic computing as a tool for automated knowledge extraction and transfer across problems in evolutionary design.

Yew-Soon Ong received the Ph.D. degree in artificial intelligence in complex design from the Computational Engineering and Design Center, University of Southampton, Southampton, U.K., in 2003.

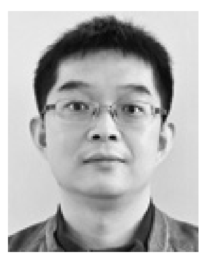
He is a Professor and the Chair of the School of Computer Science and Engineering, Nanyang Technological University, Singapore, where he is the Director of the Data Science and Artificial Intelligence Research Center, the Director of the A*Star SIMTECH-NTU Joint Laboratory on Complex Systems, and a Principal Investigator of the Data Analytics and Complex System Programme in the Rolls-Royce, NTU Corporate Laboratory. His current research interests includes computational intelligence span across memetic computation, complex design optimization, intelligent agents, and big data analytics.

Dr. Ong was a recipient of the 2015 IEEE Computational Intelligence Magazine Outstanding Paper Award and the 2012 IEEE Transactions on Evolutionary Computation Outstanding Paper Award for his research pertaining to Memetic Computation. He is the Founding Editor-in-Chief of the IEEE Transactions on Emerging Topics in Computational Intelligence, an Associate Editor of the IEEE Transactions on Evolutionary Computation, the IEEE Transactions on Neural Network and Learning Systems, and the IEEE Transactions on Cybernetics.

Kay-Chen Tan received the B.Eng. degree (First Class Hons.) in electronics and electrical engineering and the Ph.D. degree from the University of Glasgow, Glasgow, U.K., in 1994 and 1997, respectively.

He is a Full Professor with the Department of Computer Science, City University of Hong Kong, Hong Kong. He has published over 200 refereed articles and six books.

Dr. Tan is the Editor-in-Chief of the IEEE Transactions on Evolutionary Computation, was the Editor-in-Chief of the *IEEE Computational Intelligence Magazine* from 2010 to 2013, and currently serves as the Editorial Board Member of over ten journals. He is an elected member of the IEEE CIS AdCom from 2017 to 2019.



A. K. Qin (S'06–M'07–SM'12) received the B.Eng. degree from Southeast University, Nanjing, China, in 2001, and the Ph.D. degree from Nanyang Technology University, Singapore, in 2007.

From 2007 to 2012, he was with the University of Waterloo, Waterloo, ON, Canada, and with INRIA, Rocquencourt, France. Since 2013, he has been the Vice-Chancellor's Research Fellow, a Lecturer, and a Senior Lecturer with RMIT University, Melbourne, VIC, Australia. In 2017, he joined the Swinburne University of Technology, Melbourne, VIC, Australia, as an Associate Professor. He has produced over 80 publications. His current research interests include evolutionary computation, machine learning, computer vision, GPU computing, and services computing.

Dr. Qin was a recipient of the 2012 the IEEE Transaction on Evolutionary Computation Outstanding Paper Award and the Overall Best Paper Award at the 18th Asia–Pacific Symposium on Intelligent and Evolutionary Systems in 2014. One of his conference papers was nominated for the best paper award at the 2012 Genetic and Evolutionary Computation Conference in 2012. He is co-chairing the IEEE Emergent Technologies Task Forces on "Collaborative Learning and Optimization" and "Multitask Learning and Multitask Optimization."