

Deep Learning in Sheet Metal Bending With a Novel Theory-Guided Deep Neural Network

Shiming Liu, Yifan Xia, Zhusheng Shi, Hui Yu, *Senior Member, IEEE*, Zhiqiang Li, and Jianguo Lin

Abstract—Sheet metal forming technologies have been intensively studied for decades to meet the increasing demand for lightweight metal components. To surmount the springback occurring in sheet metal forming processes, numerous studies have been performed to develop compensation methods. However, for most existing methods, the development cycle is still considerably time-consuming and demands high computational or capital cost. In this paper, a novel theory-guided regularization method for training of deep neural networks (DNNs), implanted in a learning system, is introduced to learn the intrinsic relationship between the workpiece shape after springback and the required process parameter, e.g., loading stroke, in sheet metal bending processes. By directly bridging the workpiece shape to the process parameter, issues concerning springback in the process design would be circumvented. The novel regularization method utilizes the well-recognized theories in material mechanics, Swift's law, by penalizing divergence from this law throughout the network training process. The regularization is implemented by a multi-task learning network architecture, with the learning of extra tasks regularized during training. The stress-strain curve describing the material properties and the prior knowledge used to guide learning are stored in the database and the knowledge base, respectively. One can obtain the predicted loading stroke for a new workpiece shape by importing the target geometry through the user interface. In this research, the neural models were found to outperform a traditional machine learning model, support vector regression model, in experiments with different amount of training data. Through a series of studies with varying conditions of training data structure and amount, workpiece material and applied bending processes, the theory-guided DNN has been shown to achieve superior generalization and learning consistency than the data-driven DNNs, especially when only scarce and scattered experiment data are available for training which is often the case in practice. The theory-guided DNN could also be applicable to other sheet metal forming processes. It provides an alternative method for compensating springback with

significantly shorter development cycle and less capital cost and computational requirement than traditional compensation methods in sheet metal forming industry.

Index Terms—Data-driven deep learning, deep learning, deep neural network (DNN), intelligent manufacturing, machine learning, sheet metal forming, springback, theory-guided deep learning, theory-guided regularization.

I. INTRODUCTION

SHEET metal forming has been employed in industry for centuries. Among different forming techniques, sheet bending and stamping are most widely used in sheet metal forming industry [1]. These techniques have been continually developed for over a century, especially in the last couple of decades, to meet the increasing demand for lightweight metal components with high strength to reduce weight and increase fuel efficiency in the automobile and aviation industries, due to the stringent emission control. However, subject to its high yield strength-to-elastic modulus ratio, high strength aluminum alloys and steels are predisposed to notably large springback. As a consequence, the manufacturing accuracy of conventional metal forming techniques would suffer. To address this issue, vast research on estimating the springback behavior in different sheet metal forming processes has been conducted experimentally and numerically [2]–[4].

To surmount the springback in sheet metal forming process, most of the current research focuses on developing compensation methods in numerical simulations or experiments [5], [6]. With finite element (FE) simulations, Lingbeek *et al.* [7] developed the smooth displacement adjustment (SDA) method and the surface controlled overbending (SCO) method to optimize the tool shape for a deep drawing process to increase the geometry accuracy of the product after springback. Lin *et al.* [8] developed a novel advanced forming method, hot form quench (HFQ[®]), for forming complex-shaped high-strength aluminum alloy sheets, which integrates hot stamping process with heat treatment and improves geometrical accuracy by minimizing springback. Wang *et al.* [9] developed a three-dimensional (3D) scanning method to replace the traditional special fixture for inspecting the part during the stamping of a rear reinforcement plate of a door panel, which was stated to increase the qualified rate of the product to a new peak. In creep age forming (CAF), Li *et al.* [10] proposed a one-step springback compensation method for singly curved products and an accelerated method for CAF tool design with reference to the springback compensation curves, which are established based on the numerical solution of springback behavior of CAF process. The effectiveness of

Manuscript received October 20, 2020; revised November 12, 2020; accepted December 6, 2020. This work was supported by Aviation Industry Corporation of China (AVIC) Manufacturing Technology Institute (MTI) and in part by China Scholarship Council (CSC) (201908060236). Recommended by Associate Editor Xin Luo. (*Corresponding author: Zhusheng Shi.*)

Citation: S. Liu, Y. Xia, Z. Shi, H. Yu, Z. Li, and J. Lin, "Deep learning in sheet metal bending with a novel theory-guided deep neural network," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 3, pp. 565–581, Mar. 2021.

S. Liu, Z. Shi, and J. Lin are with the Department of Mechanical Engineering, Imperial College London, London, SW7 2AZ, UK (e-mail: shiming.liu17@imperial.ac.uk; zhusheng.shi@imperial.ac.uk; jianguo.lin@imperial.ac.uk).

Y. Xia and H. Yu are with the School of Creative Technologies, University of Portsmouth, Portsmouth, PO1 2DJ, UK (e-mail: yifan.xia@myport.ac.uk; hui.yu@port.ac.uk).

Z. Li is with the Manufacturing Technology Institute, AVIC, Beijing 100024, China (e-mail: zqlee98@126.com).

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/JAS.2021.1003871

the two methods was demonstrated by CAF experimental tests. Although springback occurring in different sheet metal forming processes has been significantly alleviated by the increasing amount of research in this area, the development cycle is still considerably time-consuming and demands high computational or capital cost. In this context, the impressive prediction capability and low temporal cost of machine learning technology has attracted researchers to harness this novel artificial intelligence (AI) technology in sheet metal forming industry.

Over the last two decades, machine learning technology has been vastly applied to various manufacturing processes, such as metal forming [11], semiconductor [12] and nanomaterials manufacturing [13]. In sheet metal forming industry, supervised learning technology, compared to unsupervised learning [14] and reinforcement learning [15], is most prevailing. The research focus of supervised learning in sheet metal forming processes mainly resides in classification and regression. Traditional supervised learning algorithm includes support vector machines (SVM), decision tree, Naive Bayes classifier and K-nearest neighbors (KNN) algorithm [16], [17]. Liu *et al.* [18] used SVM to predict the springback of a micro W-bending process, and the SVM was demonstrated to have high prediction accuracy and generalization performance by comparing its prediction with experimental results. Dib *et al.* [17] compared the performance of several machine learning (ML) algorithms (multilayer perceptron, random forest, decision tree, naive Bayes, SVM, KNN and logistic regression) in predicting springback and maximum thinning in a U-channel and square cup forming processes. Among these models, multilayer perceptron was reported to be the best in identifying the springback, with slightly higher score than SVM. Apart from bending, Abdessalem *et al.* [19] compared the performance of a quadratic response surface method (RSM) and two least square support vector regression models (SVRs), with polynomial kernel (PL-K) and radial basis function kernel (RBF-K), in predicting a limit state function (LSF) to determine the best surrogate model used in a probabilistic approach for the structure optimization of hydroformed sheet metals. It was found that both SVRs outperforms the RSM in learning the nonlinearity of the LSF.

Neural network learning subsumes shallow and deep learning, which are commonly distinguished by the depth of learning and the depth of networks. To date, research on shallow learning dominates the development of machine learning in sheet metal forming [20]. For example, Narayanasamy *et al.* [21] compared the performance of a multi linear regression model and a four-layer artificial neural network (ANN) in predicting the springback angle of an air bending process, from which the ANN exhibited higher prediction accuracy than the regression model. Guo *et al.* [22] developed a combination of error back propagation neural network and spline function (BPNN-Spline) to predict the springback angle in a V-die bending process, in which the BPNN took sheet metal thickness, punch radius, die radius and a material indication parameter as inputs. The proposed BPNN-Spline model was demonstrated to outperform the traditional ANN in predicting the bending angles at different punch

displacements. Viswanathan *et al.* [23] implemented a three-layer neural network to predict the stepped binder force trajectory at different punch displacement, thus realizing the control of springback in a plane strain channel forming process. Apart from the prediction of springback or forming parameters, shallow learning has also been used to substitute or reinforce the constitutive model for metallic material [24]–[26]. A three-layer neural network was developed by Jenab *et al.* [25] to predict the rheological behavior of AA5182-O sheet. Without the limitations of the mathematical function, the neural network was reported to outperform two phenomenological models, Johnson-Cook (JC) and Khan-Huang-Liang (KHL) models, in predicting the anisotropic rate-dependent behavior of AA5182-O. Li *et al.* [26] developed a machine-learning based JC plasticity model to capture the non-monotonic effect of the temperature and strain rate on the hardening response for DP800 steel. By combining shallow neural network and JC model, it was found that all experimental data can be described with high accuracy. In addition to using human designed features as the input to the neural network like above, Hartmann *et al.* [27] devised a four-layer network to predict the optimal tool path, by learning a processed form of desired workpiece geometry, for the production of sheet metals in an incremental sheet metal free-forming process. However, special attention had to be paid to the design of input and output to obtain good learning efficiency. It was claimed that, in order to cover the component spectrum, the cardinality of the training set needed to be significantly increased.

Unlike numerous studies in shallow learning, limited research has applied deep learning technology to sheet metal forming thus far. Since the great success of deep convolutional neural networks (CNNs) in learning and extracting physical features or even representations of data with multiple levels of abstraction from raw image-based inputs [28], [29], deep learning has brought breakthroughs in computer vision and image/speech/audio processing [30]–[32]. Reaping the benefits from the notable advancements in computer science technology during the last two decades, deep learning has just revealed its extraordinary learning capability in discovering the latent pattern and regularities in a given data domain. With deep learning, instead of designing hand-crafted features as the input, the workpiece geometry information is fed into the network to achieve end-to-end learning and reduce the potential bias introduced in the design of input representation. It was believed to outperform shallow neural network learning and traditional machine learning techniques, such as KNN algorithm, when handling AI-level tasks with high data dimensions and complex functions [33]. Consequently, researchers in sheet metal forming industry start resorting to deep neural network (DNN), which is an approximation model used in deep learning and consisting of relatively deep layers of neurons, for high-level learning or optimization. With deep learning, Hamouche *et al.* [34] demonstrated that a deep CNN can replace the traditional rule-based implementations in classification and selection of sheet forming processes for its higher accuracy rates. Jaremenko *et al.* [14] developed a CNN to determine the forming limit

curve by learning and clustering the common patterns existing in sheet metal geometries before defects occur. DNNs have also been applied for the manufacturability prediction of sheet metals [35] and utilized as a surrogate model (or meta-model) [36], [37]. It was seen that deep learning technology was mostly applied to classification problems in sheet metal forming processes. However, most optimization problems in engineering involves parametric regression.

Yet, overfitting is a challenge in pure data-driven deep learning research, which frequently occurs when only scarce data are available or with improper training data structures. Besides, the inferior interpretability of the normal DNN has caused confidence shrinkage in its predictions, which confers it a title of “black box”. Acquiring massive experimental data in practice is also prohibitively expensive. Consequently, physics-informed neural network (PINN) was developed to prevent overfitting problems and alleviate the low interpretability issue by respecting prior expertise and guiding the learning process with governing equations in relevant theories. The most prevailing form of physics-informed deep learning is the physics-informed data assimilation regression with physics-informed regularization method. In this method, one or several governing equations in the research area, normally nonlinear partial differential equations (PDE), are used and converted to one or several physics-informed regularization term in the objective function. In [38]–[41], PINNs were used to predict numerical solutions of governing equations in computational fluid dynamics (CFD). It was reported to outperform pure-data driven DNNs with less training data. Besides of guiding deep learning with known governing PDEs, Tartakovsky *et al.* [38] used PINN to predict the unknown state-dependent coefficient in a nonlinear diffusion equation. Yao *et al.* [42] developed an FEA (finite element analysis)-Net which embeds the governing equation used in finite element analysis into the DNN design and consists of an FEA convolution network and an inference network, to predict FE simulation results for some simplified problems. These studies have shown the superiority of PINN over pure data-driven DNNs in generalization accuracy and demand of training data.

However, no PDE-like governing equations can be used to delineate the main physical features in most Lagrangian-mechanics-like practical engineering problems with changing coordinates, especially when macro-scale parameters are of interests. To the authors’ knowledge, no attempt has been so far reported that utilizes a data-driven or theory-guided DNN (TG-DNN) in sheet metal forming industry for high-level optimization. The aim of the research is to fill in this gap. Thus, in this paper, two DNNs trained by pure training data or a novel theory-guided regularization method were developed for a learning system, named intelligent sheet metal forming system (ISMFS), and evaluated by applying them to sheet metal bending processes. By learning the intrinsic relationship between the workpiece shape after springback and the loading stroke, the springback could be compensated. The TG-DNN exploits a well-recognized theory in material mechanics, named Swift’s law, as the substitute for the PDE governing

equations to regularize the network learning. Instead of regularizing the prediction of outputs of interests, the TG-DNN incorporates multi-task learning configurations and improves the network generalization by regularizing the learning of extra tasks. With this method, the weights/biases distribution of the network would be potentially altered in a way that also facilitates the learning of the primary task, which has been corroborated in results obtained from a series of comprehensive numerical experiments in this research. A support vector regression model was also trained in this research, whose predictions were used as the baseline of learning results. The results in this research show that both data-driven and TG-DNN outperforms the SVR at the test phase. The TG-DNN can preserve high learning capability when only scarce training data are available, while the generalization accuracy of pure data-driven DNN has seen sufferings.

The contributions of this paper include: 1) filling the gap of forming process parameter prediction with deep learning in sheet metal forming industry; 2) proposing a novel theory-guided deep learning method which outperforms both traditional machine learning and data-driven deep learning methods in the prediction of non-linear behavior in sheet metal bending processes, when only scarce and scattered data are available; 3) improving the interpretability of the traditional neural model applied to metal forming industry; 4) providing an alternative method for springback compensation with shorter development cycle and less capital cost than traditional methods.

The remainder of this paper is organized as follows: Section II introduces the methodologies in this research, which include the experimental setup and FE models of sheet-metal bending experiments, the DNN architecture, theory-guided regularization using Swift’s law and data acquisition methods. Section III presents the learning results of all the models and the discussions on them. Finally, Section IV concludes the paper, with some discussions on the benefits of the TG-DNN to the practical development cycle of sheet metal forming processes and potential future works.

II. METHODOLOGY

This research presents a series of numerical experiments to evaluate the efficacy of both data-driven DNN and TG-DNN for sheet metal bending. Both DNNs are trained to predict the required loading stroke for sheet metal bending processes, given only the computer-aided design (CAD) data of the desired workpiece to be produced. This section presents the design and setups involved in the numerical experiments.

A. Bending Experiments Setup

To comprehensively evaluate the effectiveness of the data-driven DNNs and TG-DNN, two common sheet metal bending techniques, four-point bending and air bending, were employed with two commonly used material for sheet metals, AA6082 [43] and SS400 [44]. Abaqus 2019 was used to perform the numerical experiments and export the experimental data.

1) The Shape of the Components to be Manufactured

Fig. 1 presents an example of the shape of a target component and its geometry representation for four-point bending and air bending, respectively. The top surface of the target component is extracted and split into a number of uniformly distributed data points, which are then pre-processed to grayscale map before fed into DNNs as stated in Section II-B-1).

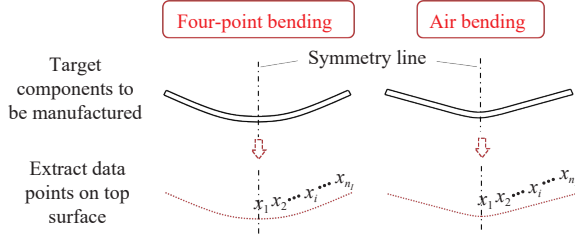


Fig. 1. Examples of target components to be manufactured using four-point bending and air bending respectively and geometry representation for assessing the error between the designed and manufactured components. Only the data points of half of the geometry were used because of the symmetric shape.

2) Numerical Procedures

Fig. 2 present the FE models and experimental setups for four-point bending and air bending. As can be seen in the FE model of the four-point bending in Fig. 2(b), the punch is only allowed to translate along Y axis and given a displacement in Y direction. The Die is fixed, and the workpiece is symmetric about its symmetry line. The friction coefficients between tools and workpiece in both cases were set to be zero for simplicity.

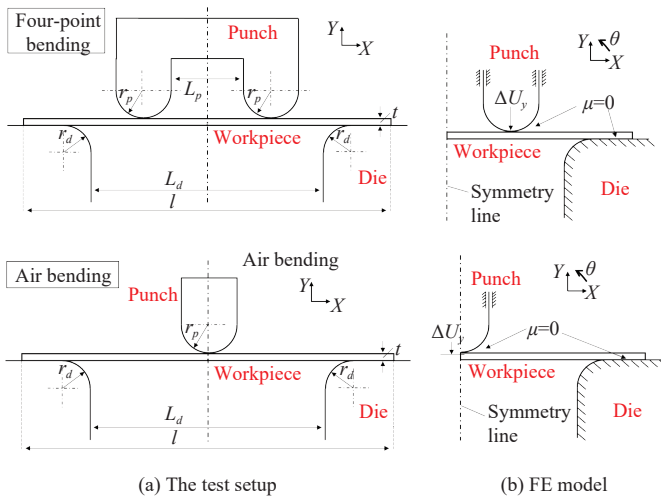


Fig. 2. The geometry, test setup and plane strain FE model for the four-point bending and air bending.

The specification of the experimental setup for four-point bending and air bending shown in Fig. 2(a) is presented in Table I.

As these numerical experiments were performed in 2-D scale, plane strain element with quadratic geometric order was used to mesh the part of workpiece. In Abaqus 2019, it is

TABLE I
EXPERIMENTAL SETUP DATA FOR FOUR-POINT BENDING AND AIR BENDING

Item	Four-point bending (mm)	Air bending (mm)
l	160	160
t	1	1
r_p	5	5
L_p	50	–
r_d	5	5
L_d	110	110

l : workpiece length; t : workpiece thickness; r_p : punch radius; L_p : punch distance; r_d : die radius; L_d : die opening.

defined as CPE8R: An 8-node biquadratic plane strain quadrilateral, reduced integration. The components modelling the punch and die working face were meshed with undeformable mesh, while the workpiece was set to be homogeneous and modelled with deformable mesh. A total element number of 21 328 was used in the FE simulation, after a grid-independence study performed to determine the most appropriate mesh size which assures reasonable simulation accuracy with low computational time.

3) Material Properties

The stress and strain curves of material of AA6082 and SS400 were extracted from [43], [44], with which the main properties of these materials were summarized and listed in Table II. The strain-hardening coefficient, n , and strength coefficient, K , were calculated by fitting the following equation to the data domain in log-scale stress-strain curve where plastic deformation occurs [45]

$$\sigma = K\varepsilon^n \quad (1)$$

where σ is the true stress and ε is the true strain.

TABLE II
VALUES OF THE PARAMETERS FOR THE MATERIALS INVESTIGATED

Item	AA6082	SS400
E (GPa)	70	210
γ (-)	0.33	0.26
ρ (g/cm ³)	2.71	7.86
n (-)	0.0731	0.1448
K (-)	400.22	708.27

E : Young's modulus; γ : Poisson's ratio; ρ : density; n : strain-hardening coefficient; K : strength coefficient.

B. Network Training in the Intelligent Sheet Metal Forming System

Fig. 3 shows the architecture of the ISMFS the DNNs were implanted, in which a database and a knowledge base were embedded. The database provides factual data to an FEA module where numerical computations are conducted to acquire the data for training of the TG-DNN. The learning of the DNN is guided by the prior knowledge or theories, which are stored in the knowledge base. After the DNN is trained to convergence, a new target workpiece geometry which is different from those in the training data is fed into the DNN

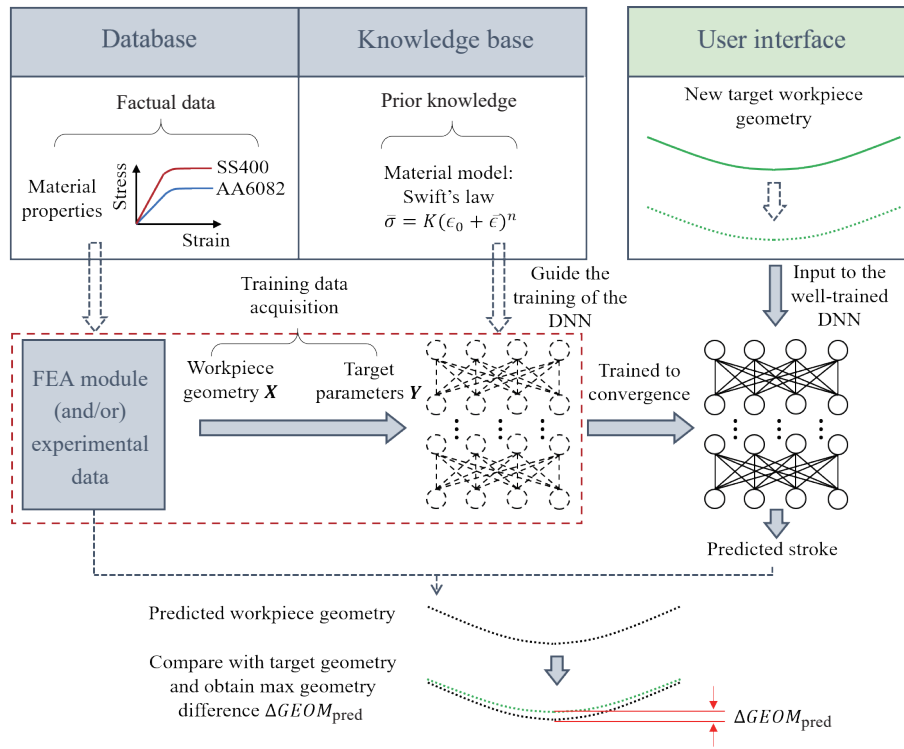


Fig. 3. The architecture of the ISMFS with the internal data flow.

through a user interface. A predicted stroke for the newly given geometry is then obtained from the DNN and used for another process simulation in the FEA module to yield the workpiece geometry after springback. Apart from obtaining the data from FEA simulations, experimental data could also be used as the training samples. In this research, all training data were obtained from FEA simulations. Finally, the data points representing the predicted and target workpiece geometry are compared to determine the maximum distance between them, which is denoted by $\Delta GEOM_{pred}$. The value of $\Delta GEOM_{pred}$ measures the prediction accuracy of the whole learning system.

1) Training Data Design and Processing

The whole dataset for training and generalization evaluation was extracted from computational experiments, conducted in the FEA module shown in Fig. 3, in which a total number of 201 simulations were performed with loading stroke ranging from 11 to 31 mm with an increment of 0.1 mm. The dataset is in a form of input-output pair, with input as the workpiece geometry representation data and output as the loading stroke. The output also includes physical parameters of true stress σ and true strain ϵ , which are acquired at the mid-point on the top surface of the workpiece before springback. Examples of the CAD of the desired workpiece after springback in four-point bending and air bending have been shown in Fig. 1. The whole dataset was then randomly split into a training set and test set. Three different training to test data ratios were used in this research: 80%/20%, 50%/50% and 20%/80%. This composition of data was designed to evaluate the performance of the TG-DNN and the data-driven DNN with different amount of training data.

Modern DNNs, especially the convolutional neural

networks (CNNs) in Computer Vision, have been proven to be able to localize and extract the important features from image-based inputs, which have been confirmed to facilitate the learning of the relationship between input data and the outputs of interests [37], [46]. With the hindsight, the input in the whole dataset was designed to be in an image-based geometry representation, known as “grayscale map”, with a 1-D array containing the local depth value about the free-end of the workpiece. Within the “grayscale map”, grayscale-values (GS), ranging from 0 to 255, were used to encode the local depth. 0 GS (“black”) denotes the location at and beyond the free-end of the workpiece (no depth), while 255 GS (“white”) denotes the location with highest depth. To further improve the learning potential of DNNs, the input data were normalized by its highest value and the outputs were standardized in the output data domain.

Fig. 4 presents some examples of inputs for the DNN in graphs and grayscale maps, respectively. It is noted that the transverse dimension in grayscale maps is used only for better visualization. There are in total 801 pixels in the longitudinal dimension, evenly distributed and aggregated to be the length of 80 mm (half length of the original workpiece blank). The map values were interpolated, with inverse distance weighting (IDW) interpolation method, from the computational results of coordinates along half of the top surface of the workpiece after springback, leveraging the symmetry of the workpiece shape.

2) Deep Neural Network

A classic convolutional neural network was designed with the concept of multi-task learning as the DNN architecture, which is schematically shown in Fig. 5. The inputs to the DNN are the greyscale maps extracted from the raw CAD data

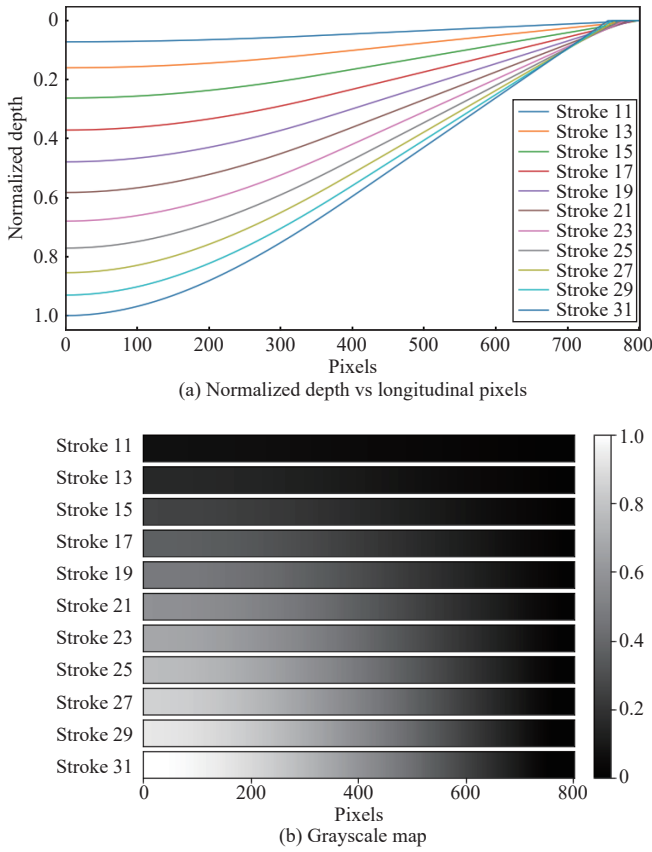


Fig. 4. Examples of workpiece geometry representations (801 pixels).

of the target workpiece after springback, and the outputs from the network include predictions of the required loading stroke and two physical parameters at the regularization spot before springback. TensorFlow Core v2.2.0 was used as the platform to establish and train the neural networks.

The convolutional layer utilizes the discrete convolution operation to shift the kernel function through the input space, which is defined as follow:

$$g^k(i^k) = (K^k * I^k)(i^k) = \sum_{t=-\infty}^{\infty} I^k[i^k - t] \cdot K^k[t] \quad (2)$$

where $\forall i^k \in [1, m^k], \mathbb{Z}$. $g^k(\cdot)$ represents the convolution function, the superscript “ k ” denotes the layer number. i^k denotes the element of the output dataset at the k th layer of network which is of dimension of m^k . $K^k(\cdot)$ and $I^k(\cdot)$ represents the kernel function and the input at k th layer. The kernel for convolutional layer is of size of 3×1 . The altering of t from $-\infty$ to ∞ signifies the shift of the kernel through the input space.

Thus, one single convolutional layer is then defined as follow:

$$\widehat{Y}_{\text{conv}}^k(\mathbf{X}^k; \theta^k) = f(g_d^k(i^k) + b_d^k) \quad (3)$$

where $\forall i^k \in [1, m^k], \mathbb{Z}; \forall d \in [1, d^k], \mathbb{Z}$. $\mathbf{X}^k = [x_1^k, \dots, x_{m^{k-1}}^k]^T$, $\mathbf{X}^k \in \mathbb{R}^{m^{k-1} \times d^{k-1}}$ are the extracted geometry features from the $(k-1)$ th layer of the network from the original workpiece geometry $\mathbf{X} = [x_1, \dots, x_{n_1}]^T$, $\mathbf{X} \in \mathbb{R}^{n_1 \times 1}$. d^k denotes the number of kernels used at the k th convolutional layer. $\widehat{Y}_{\text{conv}}^k =$

$[y_1^k, \dots, y_{m^k}^k]^T$, $\widehat{Y}_{\text{conv}}^k \in \mathbb{R}^{m^k \times d^k}$ is the output from the k th layer, which represents the newly extracted geometry features at this layer. θ^k represents all the parameters involved at the k th layer, including weights and biases. $(b_{i^k}^k) \in \mathbb{R}^{1 \times d^k}$ is the bias vector. $f(\cdot)$ is known as activation function or transfer function, rectified linear unit (ReLU) activation function is used in this DNN for regression.

The function of the max pooling layer is defined as follow:

$$\widehat{Y}_{\text{mPool}}^k(\mathbf{X}^k) = \widehat{y}_{i^k}^k = \max_{1 \leq h \leq H^k} x_{(i^k-1) \times H^k + h} \quad (4)$$

where $\forall i^k \in [1, m^k], \mathbb{Z}; \forall h \in \mathbb{Z}$. H^k denotes the height of the pooling at the k th pooling layer. The kernel for the max pooling layer is of size of 2×1 .

For a single layer of neurons of fully connected layers, its function is defined as follow:

$$\widehat{Y}_{\text{FCL}}^k(\mathbf{X}^k; \theta^k) = f\left(\sum_{i^{k-1}=1}^{m^{k-1}} x_{i^{k-1}} w_{i^{k-1}, i^k}^k + b_{i^k}^k\right) \quad (5)$$

where $\forall i^k \in [1, m^k], \mathbb{Z}; \forall i^{k-1} \in [1, m^{k-1}], \mathbb{Z}$.

Therefore, the function defining the whole DNN shown in Fig. 5 is given as follow:

$$\widehat{Y}(\mathbf{X}; \theta) = \widehat{Y}_{\text{FCL}}^{14}\left(\left(\dots \widehat{Y}_{\text{conv}}^3\left(\widehat{Y}_{\text{mPool}}^2\left(\widehat{Y}_{\text{conv}}^1(\mathbf{X}; \theta^1)\right); \theta^3\right)\dots\right); \theta^{14}\right) \quad (6)$$

where $\widehat{Y}(\mathbf{X}; \theta)$ is the prediction outputs from the DNN, which includes the predictions of the outcomes of the three learning tasks: loading stroke, true stress and true strain. As max pooling layer does not have trainable parameters, there are no θ^2 , θ^5 and θ^9 in (6). The workpiece geometry deformed using the predicted stroke is compared with the target geometry to determine maximum discrepancy, which measures the prediction accuracy of the DNN.

An objective function, also named loss function [47], is used to measure the distance between the network outputs and the true values. Mean square error (MSE), or quadratic loss, is used as the standard objective function for the training of pure data-driven DNN

$$J(\theta; \mathbf{X}, \mathbf{Y}) = \frac{1}{m^s} \sum_{j=1}^{m^s} (y_j^s - \widehat{y}_j^s)^2 \quad (7)$$

where $J(\cdot)$ is the objective function, θ represents the trainable parameters in the network including weights and biases. \mathbf{X} and \mathbf{Y} (composed of y_j^s) represent the workpiece geometry representation (grayscale map data points) and the true values of the learning tasks to manufacture the target workpiece, in which \mathbf{Y} is named the labelled output. m^s , equal to 3 in this research, is the total number of learning tasks. \widehat{y}_j^s , calculated from (6), is the predicted values of the learning tasks to manufacture the workpiece.

With backpropagate algorithm [29] and chain rule, the weights matrices and bias vectors in the network are iteratively calibrated to obtain as low outcome of the objective function as possible. The calibration goal can be expressed as following:

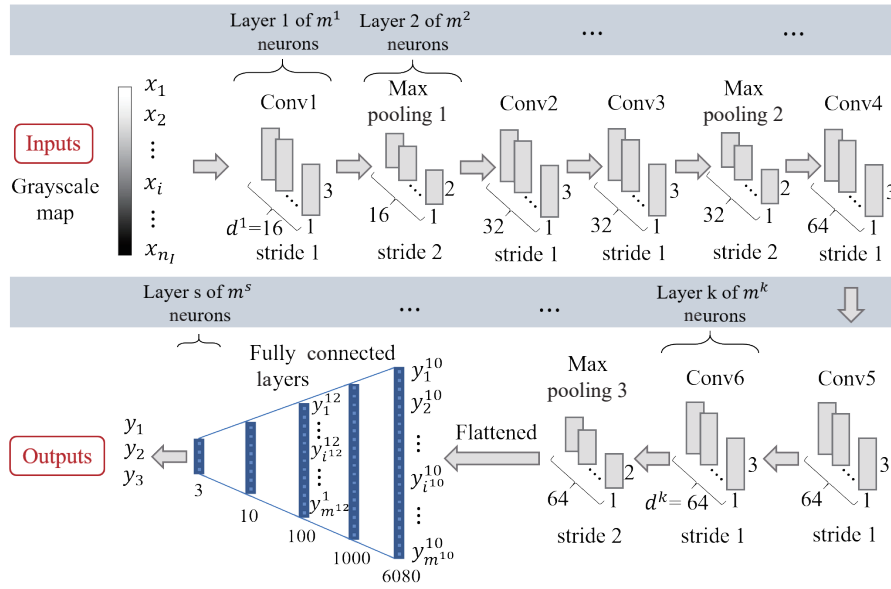


Fig. 5. Architecture of the deep neural network in the ISMFS.

$$\theta^* = \arg \min_{\theta} J(\theta; \mathbf{X}, \mathbf{Y}) \quad (8)$$

where θ^* are the optimal θ to obtain the minimum value of the objective function at the end of the optimization. Several prevailing optimization algorithms have been developed to realize this calibration process over the last decades, with their different upsides and downsides. The commonly used learning algorithms include stochastic gradient descent (SGD) [33], Adam [48] and root mean square propagation (RMSProp) [49].

3) Parameter Norm Regularization

Regularization methods have been applied to traditional machine learning for decades, and they are proved to be also effective in limiting the capacity of DNNs after the advent of deep learning. Parameter norm regularization introduces a parameter norm penalty, $\alpha\Omega(\theta)$, to the original objective function shown in (7) and obtain a regularized objective function [33]

$$\hat{J}(\Theta; \mathbf{X}, \mathbf{Y}) = J(\theta; \mathbf{X}, \mathbf{Y}) + \alpha\Omega(\theta) \quad (9)$$

where $\alpha \in [0, \infty)$ is a hyperparameter used to control the relative contribution of the norm penalty term, $\Omega(\theta)$, regarding to the original objective function, $J(\theta; \mathbf{X}, \mathbf{Y})$. Θ represents all the parameters involved in this network, including weights, biases and hyperparameters. It can be seen that the penalty term is a function of trainable parameters, θ . By controlling the training modes of the network parameters, the overfitting issues in many research studies can be addressed or alleviated [33].

Theory-guided deep neural network respects prior treasurable theoretical or empirical knowledge by leveraging the theories to regularize the learning process. Compared to pure data-driven DNN model, a well-trained TG-DNN would be more than just a “black box” as theoretical knowledge is taken into account and the credibility of its prediction results is thus increased. Similar to the general form of the regularized objective function in (9), the theory-guided loss function is regularized by theoretical laws as shown as follow:

$$\hat{J}_{TH}(\Theta; \mathbf{X}, \mathbf{Y}) = J(\theta; \mathbf{X}, \mathbf{Y}) + \alpha_{TH} J_{TH}(\theta; \mathbf{X}) \quad (10)$$

where $\alpha_{TH} \in [0, \infty)$ is a hyperparameter used to control the relative contribution of the theory-guided penalty term, $J_{TH}(\theta; \mathbf{X})$, which is defined as

$$J_{TH}(\theta; \mathbf{X}) = \frac{1}{n_{\mathcal{H}}} \sum_{i=1}^{n_{\mathcal{H}}} [\mathcal{H}(X_i, \hat{\mathbf{Y}}(X_i; \Theta))]^2 \quad (11)$$

where $\mathcal{H}(X_i, \hat{\mathbf{Y}}(X_i; \Theta))$ is a function of the inputs and outputs of the TG-DNN at one regularization spot, which measures the divergence of the theoretical laws at this spot. $n_{\mathcal{H}}$ is the total number of regularization spot. By minimizing both the loss between network outputs and labels and the divergence of the solutions from the theoretical laws, the network would be able to learn the underlying physics of the theories and generate outputs respecting these theories.

As stated in the introduction, the most common form of $\mathcal{H}(X_i, \hat{\mathbf{Y}}(X_i; \Theta))$ for PINN, to date, is the governing equation in the research area, which is normally one or several nonlinear partial differential equations (PDE). The convergence of the regularization term would guide the learning process by forcing the governing equation to tend to hold between the inputs and outputs of the PINN. In this way, the PINN could learn the underlying theoretical laws and its prediction would be more interpretable than pure data-driven DNNs. However, due to the intrinsic complexity of Lagrangian mechanics in most practical engineering problems, the material model, Swift’s law, is used as the substitute for the PDE governing equations to regularize the network learning. By implementing this theory in training, the network could learn in a way respecting the material model. A multi-task learning network configuration was used to implement the regularization. This method uses a less hard theory-guided penalty term $J_{TH}(\theta; \mathbf{X})$ than the mainstream PINNs to regularize the learning of extra tasks to improve the learning of primary ones, instead of bridging the inputs and outputs with governing equations. Multi-task learning has been employed in numerous researches for its improved

learning speed and accuracy and more chances for harder tasks to be learned than single-task learning [50], such as face detection and gender recognition in computer vision [51], natural language processing [31], speech recognition [32] and neurodegenerative disease diagnosis [52]. The rationale behind is that the learner with multi-task learning can learn related work at the same time and use these tasks as inductive bias for each other, which thus allows the domain regularities to be better learned [33].

Swift's law is a generalized power law for expressing the material-hardening behavior when plastic deformation occurs in a material [53], which has been introduced in (1). To replicate material model obtained in the one-dimensional tension test, a plasticity model is defined using a scalar equivalent stress and an equivalent plastic strain to replace the true stress and true stress in (1). The von Mises stress, σ_v , and the corresponding equivalent plastic strain, ε_{eq} , are used in this study

$$\sigma_v = K\varepsilon_{eq}^n. \quad (12)$$

Thus, (12) is the theoretical equation used in this study to regularize the learning of the TG-DNN. Two extra tasks on learning additional physical parameters (true stress and true strain) are added to the TG-DNN for regularization.

Consequently, (12) can be written into the form of $\mathcal{H}(X_i, \widehat{Y}(X_i; \Theta))$

$$\mathcal{H}(\widehat{Y}(X_i; \Theta)) = \widehat{\sigma}_v(X_i; \Theta) - K(\widehat{\varepsilon}_{eq}(X_i; \Theta))^n. \quad (13)$$

Then the theory-guided objective function regularized by Swift's law is given as follow:

$$\begin{aligned} \widehat{J}_{TH}(\Theta; X, Y) = & J(\theta; X, Y) \\ & + \frac{\alpha_{TH}}{n_{\mathcal{H}}} \sum_{i=1}^{n_{\mathcal{H}}} \left(\widehat{\sigma}_v(X_i; \Theta) - K|\widehat{\varepsilon}_{eq}(X_i; \Theta)|^n \right)^2 \end{aligned} \quad (14)$$

where an absolute value of the term $\widehat{\varepsilon}_{eq}(X_i; \Theta)$ is used to prevent the base of the exponent from being negative values, as n is less than 1.

The effectiveness of the introduced theory-guided penalty term, $J_{TH}(\theta; X)$, is manually checked for the training examples by comparing the calculated values of the penalty term and the ideal value (zero). The calculation check was performed for the whole 201 data in the four experiments shown in Section III, with the relative error to the ideal value presented in Fig. 6. It can be seen that the relative error for each training dataset is negligibly small, thus, the application of the theory-guided penalty to regularize the learning of the training data domain is theoretically valid.

4) Training Methods

For training the SVR, the radial basis function kernel was used, and the hyperparameters of “ ϵ ” and “ C ” were set as 0.1 and 1, respectively. The SVR was established and trained with scikit-learn library in Python. The Adam optimization algorithm [48] was used to train the DNNs, with default parameters of its hyperparameters ($\beta_1, \beta_2, \epsilon$) in TensorFlow Core v2.2.0. The DNNs were trained with mini-batch training method [33]. At training, the training data were shuffled after

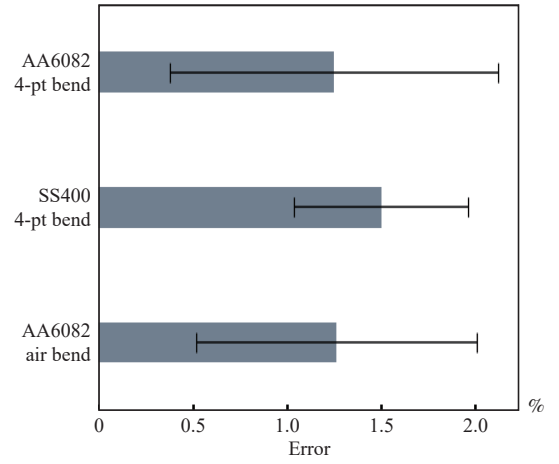


Fig. 6. The relative error between the calculated results of the theory-guided penalty term for the whole 201 examples and the ideal value (zero) for the indicated materials and processes. 4-pt bend denotes four-point bending process.

each training epoch, which indicates one full cycle through the whole given training dataset, to reduce the sequence bias to the learning. The sizes of mini-batch used in different experiments are presented in Section III.

An exponentially decaying learning rate was used in this research, as lower learning rate is recommended as the training progresses. The learning rate decaying scheme is defined as follow:

$$\eta = \eta_0 \times s_d^{\frac{T_d}{t_d}} \quad (15)$$

where η denotes the learning rate at the current decaying time step whose number is t_d , η_0 is the initial learning rate, s_d is the decaying rate and T_d is the total decaying step number. s_d and T_d were set as 0.96 and 100 000 in this research.

The hyperparameters search space for the DNNs studied are presented in Table III. The highest test accuracy that the model can reach in these search domains would be regarded as the best performance the model can achieve.

TABLE III
HYPERPARAMETERS SEARCH SPACE

	Data-driven DNN	Theory-guided DNN
Learning rate	$\eta_0 \in [10^{-8}, 10^{-3}]$	$\eta_0 \in [10^{-8}, 10^{-3}]$
Regularization coefficient	–	$\alpha_{TH} \in [10^{-4}, 1]$

C. Data Acquisition and Training Process

Fig. 7 shows the training data acquisition through the FE simulation and how the data were used for TG-DNN training, which is taking place in the ISMFS as shown by the red dashed square in Fig. 3. The experimental results of loading stroke (d_s) and two physical parameters (σ_v and ε_{eq}) at the mid-point of the workpiece top surface were collected before the springback of the workpiece. Thus, only one regularization spot was set. These data were used as the target data in calculating the physics-guided objective function as shown in

(14). The top surface geometry feature of the workpiece after springback was extracted and pre-processed to grayscale map before it was fed into the network as inputs.

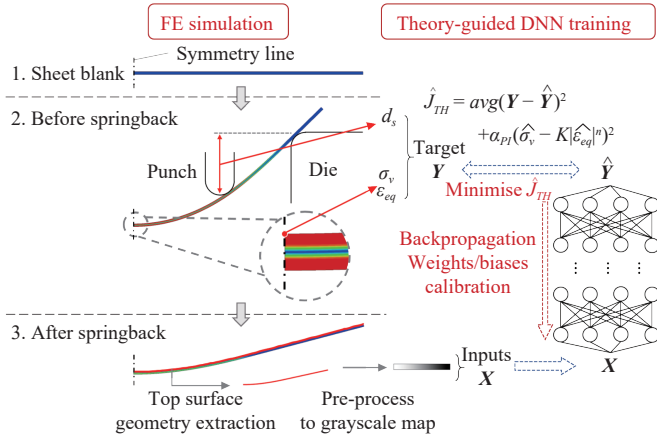


Fig. 7. Training data acquisition and data flow between FE simulation and TG-DNN training.

Fig. 8 shows the iterations in the whole learning process to obtain the optimal loading stroke to manufacture the target workpiece. The difference between the geometry of the target workpiece and the workpiece manufacture with the predicted stroke, $\Delta GEOM$, is evaluated in each iteration to update the parameters of the DNN. $|\Delta GEOM|$ is iteratively reduced through this learning process and would reach a minimum value in the end of the learning.

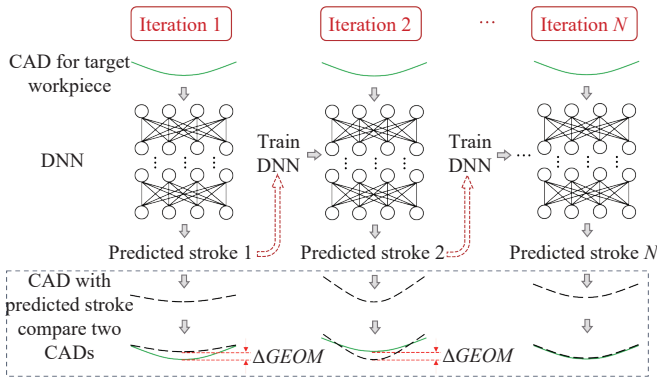


Fig. 8. The learning process to obtain the optimal loading stroke prediction.

III. LEARNING RESULTS AND DISCUSSION

All the neural models were trained with the hyperparameters search space shown in Table III. The model that result in the lowest relative L_2 norm on the test set was then trained for another 3 times, and the generalization error of the network was obtained by averaging the performance of the three trained models. The average results of $\Delta GEOM_{pred}$ from the three training is used to measure the performance of the machine learning models.

A. Comparison Between Support Vector Regression and Neural Networks

To evaluate the performance of the machine learning

models described in the methodology section (SVR, data-driven DNN and TG-DNN), the original 201 data were split into three different training-test data composition: 80%/20%, 50%/50% and 20%/80%. The training data were randomly sampled from the whole dataset for 10 times to train the models, of which the average results are shown in Fig. 9. As the theory-guided learning exploits the concept of multi-task learning, to ascertain whether this method improves learning performance, two data-driven DNNs, multi-task DNN (MT-DNN) and single-task DNN (ST-DNN), were established for comparison. The four DNNs used the same architecture as shown in Fig. 5, except for the ST-DNN which only has one neuron in the output layer. All the machine learning models were applied to predict the loading stroke, based on the target workpiece geometry, in a four-point bending process of a sheet metal with material of AA6082. The number of trainable parameters for ST-DNN is 6 217 769, and that for MT-DNN and TG-DNN is 6 217 791.

From Fig. 9, the prediction accuracy from SVR has apparent deterioration when the amount of training data is reduced, while those from the three neural models have negligible change. Furthermore, the predictions from DNNs conforms well with the true data in all three cases, while the prediction from SVR has clear deviation from its true data, especially when only 20% data were used for training. The results of the maximum distance between the target and predicted workpiece shape ($\Delta GEOM_{pred}$), which is used to measure the prediction accuracy, for the four machine learning models are shown in Table IV. From the results in the table, the data-driven models and the theory-guided model show indiscernibly good performance in all three experiments. To further evaluate the learning capability of the data-driven DNN and the TG-DNN, an experiment was performed with its results shown in the sequel.

B. Comparison Between Data-Driven DNN and Theory-Guided DNN

Because of the nearly equal performances of the data-driven and TG-DNN, another experiment was designed with new composition of training and test data. The ratio of training to test data is kept to the poorest condition as in last experiment (about 20%/80%), while the training data were chosen to eschew large continuous regions of data domain which were unseen at training, instead of having a relatively uniform data for training obtained from random sampling. Two data compositions, named symmetric (includes stroke values in the range of 11–12 mm, 21–22 mm and 30–31 mm) and asymmetric training data (includes stroke values in the range of 11–15 mm and 30–31 mm), were designed in this study. This design of training data is to better reflect practical situations in industry, where experimental data are expensive and scattered. In addition, to evaluate the learning consistency of neural models, two more experiments were designed to apply both data-driven DNNs and TG-DNN to 1) four-point bending process with a new material of SS400 and 2) air bending process with the same material of AA6082. Thus, the learning capabilities of the neural models were evaluated in

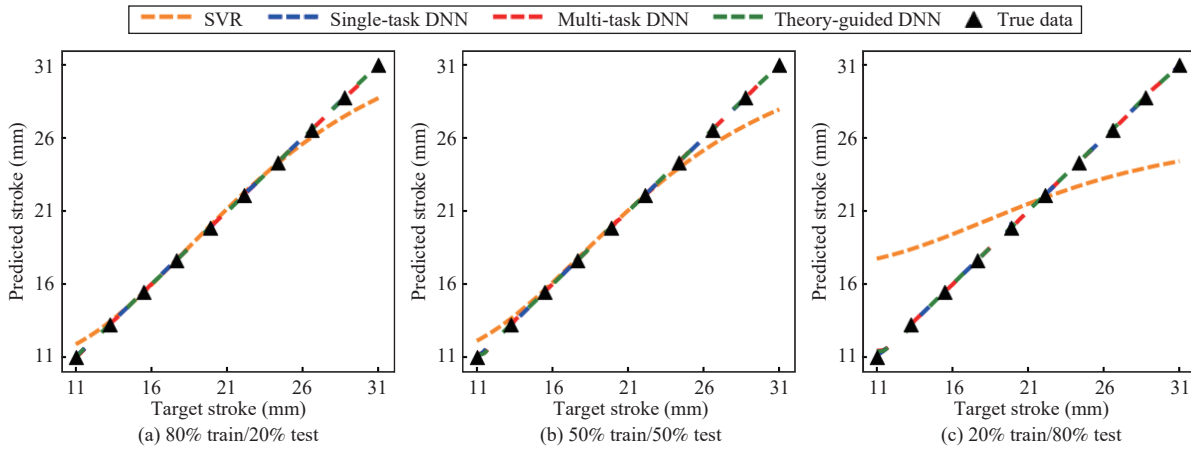


Fig. 9. Comparison between the true data and predictions of the data domain from SVR, ST-DNN, MT-DNN, and TG-DNN, which were trained with data of three different ratios of training to test data. The true data were obtained from four-point bending test of AA6082. The training data were randomly sampled from the whole dataset.

TABLE IV
COMPARISON OF THE AVERAGE OF THE MAXIMUM DISTANCE BETWEEN THE TARGET AND PREDICTED WORKPIECE SHAPE ($\Delta GEOM_{pred}$) FROM THE WELL-TRAINED SVR AND DNNs EVALUATED ON THE TOTAL DATA DOMAIN (MM)

Data composition	80% train/ 20% test	50% train/ 50% test	20% train/ 80% test
SVR	0.43 ± 0.05	0.68 ± 0.04	3.21 ± 0.24
Single-task DNN	0.01 ± 0.01	0.02 ± 0.01	0.03 ± 0.01
Multi-task DNN	0.02 ± 0.01	0.03 ± 0.02	0.03 ± 0.01
Theory-guided DNN	0.02 ± 0.01	0.02 ± 0.01	0.03 ± 0.01

experiments with three varying conditions:

1) Different raw data structure: using symmetric/asymmetric training data, which were acquired from the four-point bending process of the AA6082 workpiece.

2) Different workpiece material: using symmetric training data, which were acquired from the four-point bending process of the SS400 workpiece.

3) Different forming process (applications): Using symmetric training data, which were acquired from the air bending process of the AA6082 workpiece.

Fig. 10 shows the training loss curves for the three DNNs during training and their generalization performance under four different conditions. The generalization performance is measured by evaluating the well-trained DNN on the test set, and the predictions for the whole data domain from different models are plotted. It was seen that the training losses of DNNs in different cases all converge to a small value at the end of training. Due to an extra theory-guided regularization term in the objective function for the TG-DNN, the convergence of which would conflict that of the standard objective function, the converged training loss of the TG-DNN are larger than those of the pure data-driven DNNs. From the prediction curves of the three models, compared to the performance of the DNNs trained with randomly sampled training data in Fig. 9(a), the prediction from the TG-DNN remains a good fitting to its target value, while the data-driven DNNs trained with scattered data in this section fail to learn

the true function in the unsupervised data domain in most experiments. Besides, the TG-DNN shows a good learning consistency from its consistent performance under various training conditions and in different applications. Table V shows the average results of $\Delta GEOM_{pred}$ from the DNNs in all experiments, in which the TG-DNN is superior to the two data-driven DNNs. It is worth noting that the MT-DNN outperforms the ST-DNN in most cases, which indicates that the learning of loading stroke could be improved by learning extra tasks of local true stress and strain.

C. Learning With Scarce Training Data

In industry, experimental data can be rather expensive, for which of great significance is the predictive model to have a good performance with very limited training data. Thus, in this section, a new experiment was designed to have a total number of 21 examples of dataset, with loading stroke ranging from 11 to 31 mm with unit increment. The whole dataset was then split into a training set and test set, with data pairs with odd numbers of loading stroke (i.e., 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31 mm) as training set and data pairs with even numbers of loading stroke (i.e., 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 mm) as test set. Thus, there are in total 11 and 10 data examples in training set and test set, respectively. To comprehensively evaluate the performance of the DNNs with scarce data, three different amounts of training data were used for training the neural networks. The original training set with 11 examples were designed to be the “full training set”. Likewise, the training data with 6 (with stroke values of 11, 15, 19, 23, 27, 31 mm) and 3 (with stroke values of 11, 21, 31 mm) evenly distributed examples are referred to as the “half training set” and “1/4 training set”.

All the networks were trained to convergence with the batch size of 2. Fig. 11 presents the generalization results of the ST-DNN, MT-DNN and the TG-DNN, well-trained with full and half original training data, on the test set. The training data were obtained from the four-point bending process of AA6082 workpiece. It can be seen that, with full and half amount of

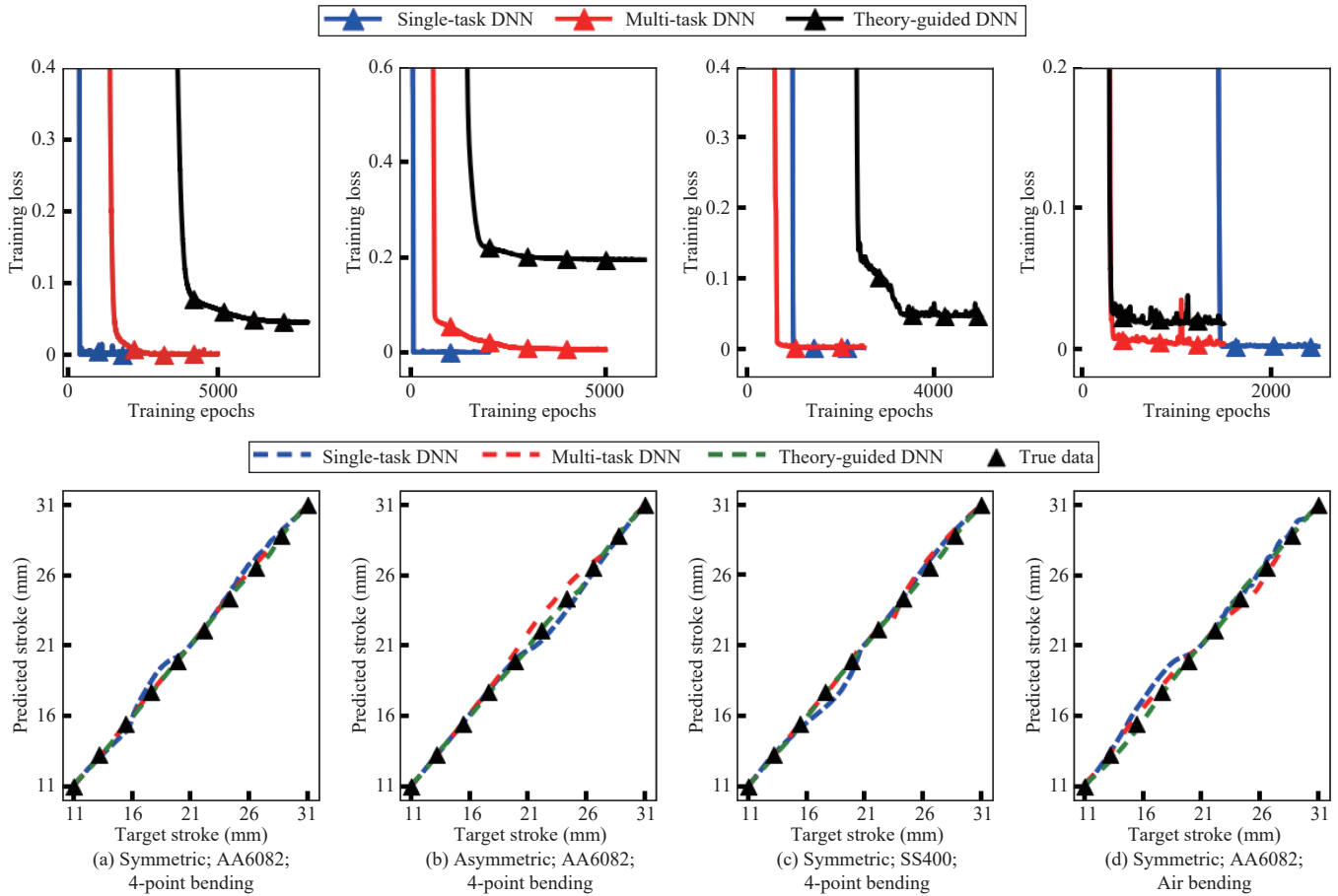


Fig. 10. The training loss curves for the three DNNs (upper row) and the evaluations of the DNNs on the total data domain (lower row) after training with about 20% training data acquired from different applications. The training data was designed to be symmetric and asymmetric to have large continuous regions of missing data domain.

TABLE V
COMPARISON OF THE AVERAGE OF THE MAXIMUM DISTANCE BETWEEN THE TARGET AND PREDICTED WORKPIECE SHAPE ($\Delta GEOM_{\text{pred}}$) FROM THE WELL-TRAINED DNNs EVALUATED ON THE TOTAL DATA DOMAIN IN ALL EXPERIMENTS (MM)

Data str	Workpiece material	Mfg process	ST-DNN	MT-DNN	TG-DNN
Sym	AA6082	4-point bending	0.34 ± 0.03	0.09 ± 0.02	0.08 ± 0.02
Asym	AA6082	4-point bending	0.26 ± 0.03	0.35 ± 0.04	0.11 ± 0.02
Sym	SS400	4-point bending	0.35 ± 0.04	0.22 ± 0.03	0.07 ± 0.01
Sym	AA6082	Air bending	0.44 ± 0.04	0.30 ± 0.05	0.20 ± 0.03

Sym: Symmetric; Asym: Asymmetric; str: structure; Mfg: Manufacturing.

training data, all DNNs perform nearly equally well. This can also be observed in Table VI, in which the values of $\Delta GEOM_{\text{pred}}$ between the target and predicted workpiece geometry from all models are negligibly small with full and half training data.

With 1/4 training samples, the training data were also designed to be structurally symmetric (includes stroke values of 11, 21 and 31 mm) and asymmetric (includes stroke values of 11, 13 and 31 mm) for four-point bending of AA6082 like in Section III-B. The models were also applied to applications of 1) four-point bending process with a new material of SS400 and 2) air bending process with the same material of AA6082 to comprehensively evaluate the performance and learning

consistency of the data-driven DNNs and TG-DNN.

Fig. 12 shows training loss curves for all the DNNs during training and the generalization performance of them under four different conditions. The generalization performance is measured by evaluating the well-trained DNN on the test set (with stroke values of 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 mm). It was seen that the training losses of all the DNNs in different cases all converge to a small value at the end of training. Similar to the results in experiments with 201 total data shown in Fig. 10, the TG-DNN has a higher converged loss than the data-driven DNNs due to its extra theory-guided regularization term in the objective function. From the prediction graphs, the predictions of stroke values on the test

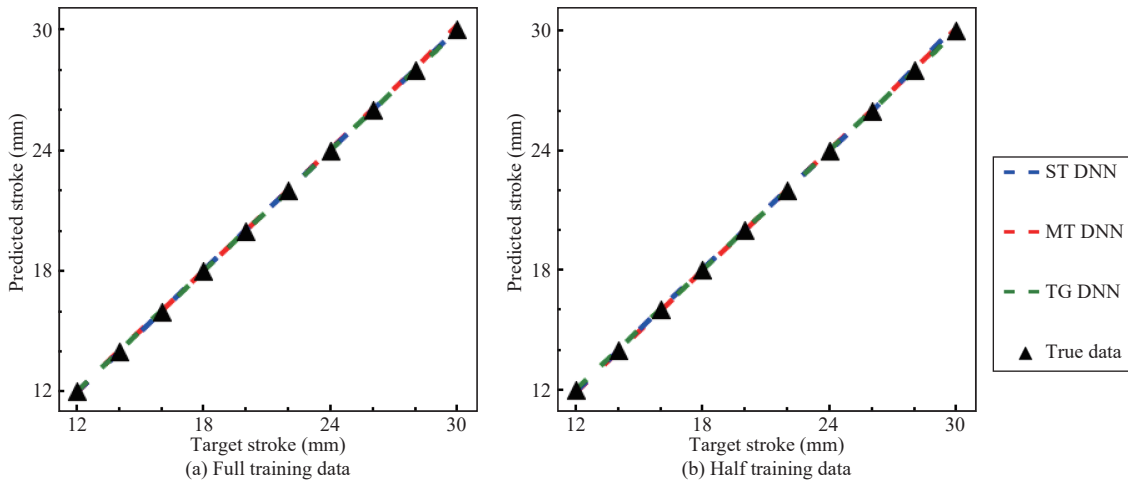


Fig. 11. Comparison of generalization capability on test set between the data-driven and theory-guided DNNs with (a) full data; (b) half data, obtained from four-point bending process of AA6082 workpiece.

TABLE VI

COMPARISON OF THE AVERAGE OF THE MAXIMUM DISTANCE BETWEEN THE TARGET AND PREDICTED WORKPIECE SHAPE ($\Delta GEOM_{\text{pred}}$) FROM THE WELL-TRAINED DATA-DRIVEN AND THEORY-GUIDED DNNs EVALUATED ON THE TEST SET. THE DATA WERE OBTAINED FROM FOUR-POINT BENDING PROCESS OF AA6082 WORKPIECE (MM)

	ST-DNN		MT-DNN		TG-DNN	
Training data	Full data	Half data	Full data	Half data	Full data	Half data
Test accuracy	0.03 ± 0.01	0.08 ± 0.01	0.05 ± 0.01	0.08 ± 0.01	0.03 ± 0.01	0.10 ± 0.01

set from the TG-DNN have a good match to the target values in all cases. Compared with the multi-task and single-task learning results, where generalization errors occur evidently, the TG-DNN exhibits effective learning and still holds high generalization performance in the data domain away from the training set. With only 1/4 training data, the data-driven DNN fails to learn the true function in the unsupervised data domain, while the TG-DNN succeeds to retain high learning capability. From the average results of $\Delta GEOM_{\text{pred}}$ in different applications shown in Table VII, with 1/4 training data, the results of $\Delta GEOM_{\text{pred}}$ measured from TG-DNN only have an increase less than 0.1 mm than those obtained from training with full and half data shown in Table VI. Thus, the TG-DNN exhibits a significantly more robust and consistent learning capability than data-driven DNNs when applied to problems with different data structures and amount of the training data, different materials of the workpiece and different bending processes.

It is worth noting that, in the case with asymmetric training data shown in Fig. 12(b), the TG-DNN also managed to learn the true function in spite of the considerably increased learning difficulties due to large amount of missing data domain for training purpose (between target stroke of 13 and 31 mm). The predictions from both MT-DNN and ST-DNN have evident deviation from the true function in this region. It is well known that the structure of the training dataset has significant influence on the training of neural networks, while the data structure is commonly irregular when acquiring off-the-shelf experimental data. Thus, the introduced TG-DNN, in practice, can effectively alleviate the demand on the amount

and structure of expensive experimental data, which could reduce the effort in experiment design. Also, from the prediction graphs in Figs. 10 and 12, the performance of the data-driven DNNs trained with only three examples in Fig. 12 is inferior to that of the DNNs trained with 30–50 data in Fig. 10, while the TG-DNN shows a more consistent performance. This indicates the superior learning capability of the TG-DNN than the data-driven DNN trained with very scarce data. It should be noted that other efforts have also been made to deal with missing data values, such as non-negative latent factor model [54], [55], which outperforms the state-of-the-art predictors.

Taking case 1 and case 4 (Figs. 12(a) and 12(d)) as examples, with the predicted values for target stroke of 24 mm from the DNNs, the obtained workpiece geometries deformed by four-point bending and air bending are compared with their targets, as shown in Fig. 13. It can be seen that the geometry predicted from TG-DNN is evidently closer to the target than those from ST-DNN and MT-DNN. Fig. 13 also shows how the $\Delta GEOM_{\text{pred}}$ is measured for different DNNs.

To ascertain whether the generalization accuracy of the DNN is promoted by the proposed theory-guided regularization method using Swift's law, the final training loss and test loss of the physical parameters (σ_v and ε_{eq}) from the well-trained TG-DNN and the MT-DNN in all experiments are compared as shown in Fig. 14. It can be seen that the training loss of these parameters from data-driven DNN are extremely smaller than those from the TG-DNN, which is due to the convergence conflict between the standard term and the theory-guided term in the objective function for the TG-DNN.

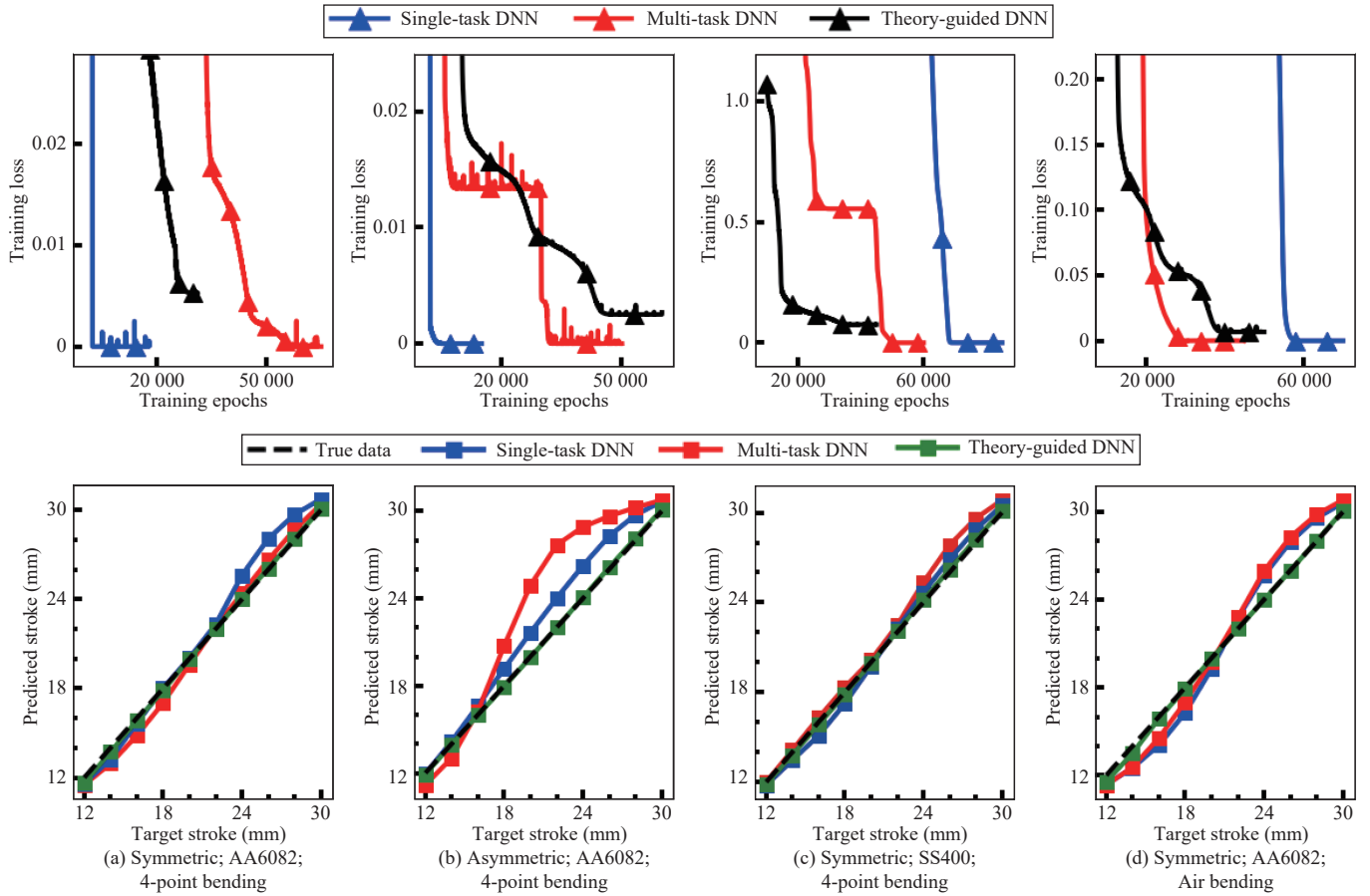


Fig. 12. The training loss curves for the data-driven and theory-guided DNNs (upper row) and the evaluations of all the DNNs on the test set (lower row) after training with 1/4 training data acquired from different applications.

TABLE VII

COMPARISON OF THE AVERAGE OF THE MAXIMUM DISTANCE BETWEEN THE TARGET AND PREDICTED WORKPIECE SHAPE ($\Delta GEOM_{\text{PRED}}$) FROM THE WELL-TRAINED DATA-DRIVEN DNNs AND THEORY-GUIDED DNN EVALUATED ON THE TEST SET IN ALL EXPERIMENTS (MM)

Data str	Workpiece material	Mfgprocess	ST-DNN	MT-DNN	TG-DNN
Sym	AA6082	4-point bending	0.80 ± 0.07	0.63 ± 0.06	0.18 ± 0.02
Asym	AA6082	4-point bending	1.26 ± 0.10	2.69 ± 0.15	0.16 ± 0.02
Sym	SS400	4-point bending	0.58 ± 0.06	0.69 ± 0.06	0.17 ± 0.01
Sym	AA6082	Air bending	1.30 ± 0.09	1.24 ± 0.07	0.19 ± 0.03

Sym: Symmetric; Asym: Asymmetric; str: structure; Mfg: Manufacturing.

However, the loss of the physical parameters on the test set from the TG-DNN are smaller than those from the data-driven DNN, which indicates its better learning outcome in the physical-parameters data domain. This corroborates the learning improvements introduced by the TG-DNN, as the physical parameters were directly regularized in the theory-guided objective function. Consequently, the superior performance of the TG-DNN than the data-driven DNN is demonstrated by the introduced theory-respecting regularization method. With the effective implementation of Swift's law for training regularization, the interpretability of the DNN is improved.

In summary, the ISMFS with the novel TG-DNN proposed in this research has shown many advantages. The successful

application of TG-DNN to sheet metal bending processes also indicates its applicability to other sheet metal forming techniques, where the physical parameters (σ_v and ε_{eq}) are, in general, monotonically related to the forming parameters. Despite of its versatility, TG-DNN could also have limitation. As the learning task of interest is not directly related to the inputs to the model with a hard theory-respecting regularization function like those in [38]–[41] due to the inherent complexity of sheet metal bending engineering problems, a more tricky process of hyperparameters tuning than those for PINN could potentially be needed. As a consequence, the method would increase the computational cost compared to the PINN with PDE governing equations. However, because of the infeasible application of PINN in this

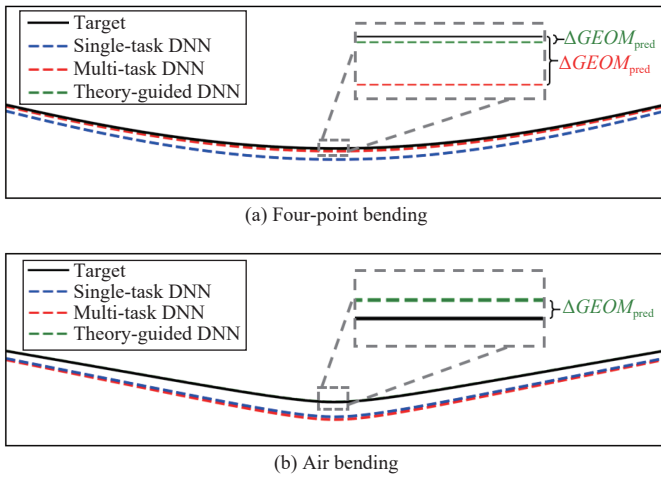


Fig. 13. Comparison between the shape of the target workpiece (with target stroke of 24 mm) and the workpiece manufactured with the stroke predicted by the data-driven and theory-guided DNNs, which were trained with 1/4 symmetric data acquired from mid-point of the AA6082 workpiece from (a) four-point bending and (b) air bending.

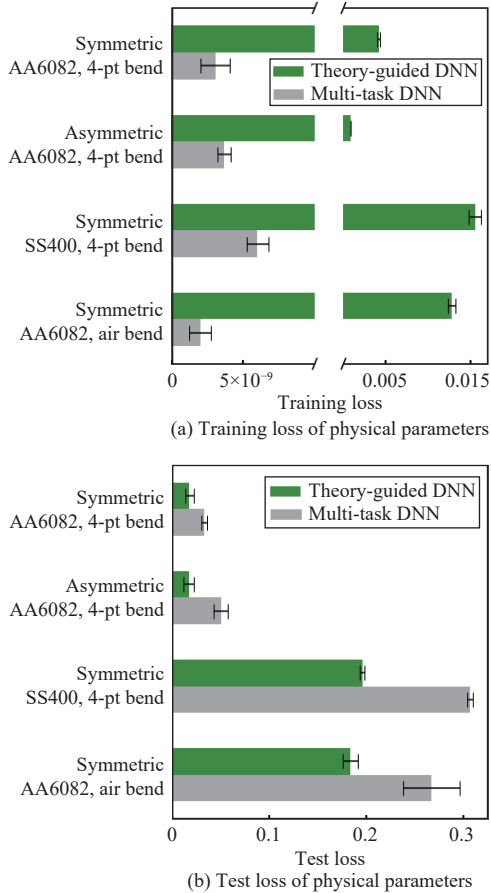


Fig. 14. Comparison between the training and test loss of physical parameters (σ_v and ε_{eq}) between well-trained MT-DNN and TG-DNN with 1/4 symmetric training data acquired from mid-point of the AA6082 workpiece from 4-point bending.

research, discussion on these two neural networks could be futile.

IV. CONCLUSIONS

In this research, a novel theory-guided regularization method for DNN training, embedded in a learning system, has been proposed. The TG-DNN, which utilizes Swift's law as the guidance, improves the learning performance by regularizing the learning of extra tasks. A series of studies have been performed to evaluate the machine learning models, including a support vector regression model, two data-driven DNNs and a TG-DNN, in learning the intrinsic relationship between the deformed workpiece geometry after springback and the corresponding forming parameter of loading stroke. The following conclusions can be drawn:

1) The neural networks, either trained in data-driven or theory-guided, exhibit blatantly better learning capabilities than a traditional machine learning model, SVR, in predicting the loading stroke of a four-point bending and an air bending process. With less training data sampled randomly, the performance of the SVR deteriorates sharply, while those of the neural models have negligible decrease.

2) The proposed TG-DNN outperforms the conventional pure data-driven DNN for its superior generalization accuracy, when only scarce and scattered experimental data are available for training. When decreasing the amount of training data to a very small value (from 30–50 to 3 in this research), the performance of the data-driven DNNs suffers while that of the TG-DNN retains consistent.

3) With varying experimental conditions, including the structure and amount of training data, workpiece material and sheet metal forming application, the TG-DNN exhibits a more robust learning capability and learning consistency than the data-driven DNNs.

4) Other than bending, The TG-DNN could also be applicable to the sheet metal forming processes where the physical parameters (σ_v and ε_{eq}) are, in general, monotonically related to the forming parameters. The theory-guided learning also improves the interpretability of the traditional data-driven DNNs.

5) Consequently, in practice, the proposed theory-guided regularization method could potentially relieve the high demand on the amount and structure of expensive experimental data. Furthermore, by implementing the theory-guided deep learning technology, the problems of considering springback in the traditional design of sheet metal bending processes could be circumvented. The TG-DNN, thus, provides an alternative method for compensating springback with significantly shorter development cycle and less capital cost and computational requirement than traditional compensation methods which are based on FE simulations and experiments.

One of the future works is to apply this TG-DNN to other sheet metal forming processes. Developing a harder regularization method, by introducing new or more theories for sheet metal bending process and/or increasing the regularization spots in the theory-guided objective function, could also be worthy of investigation. Also, the latent

relationship between the process parameters and the workpiece geometry after springback could be learned by developing equations, which could be transformed from equations describing the workpiece-tools topology before springback by adding a to-be-learned term relating the workpiece geometry before and after springback. Other approaches could be used to enhance the theory-guided model, with more advanced optimization methods [56], [57], or discrete-time delayed neural networks [58], [59].

REFERENCES

- [1] K. L. Zheng, D. J. Politis, L. L. Wang, and J. G. Lin, "A review on forming techniques for manufacturing lightweight complex-shaped aluminium panel components," *Int. J. Light. Mater. Manuf.*, vol. 1, no. 2, pp. 55–80, Jun. 2018.
- [2] L. H. Zhan, J. G. Lin, T. A. Dean, and M. H. Huang, "Experimental studies and constitutive modelling of the hardening of aluminium alloy 7055 under creep age forming conditions," *Int. J. Mech. Sci.*, vol. 53, no. 8, pp. 595–605, Aug. 2011.
- [3] W. L. Xu, C. H. Ma, C. H. Li, and W. J. Feng, "Sensitive factors in springback simulation for sheet metal forming," *J. Mater. Process. Technol.*, vol. 151, no. 1–3, pp. 217–222, Sep. 2004.
- [4] K. P. Li, W. P. Carden, and R. H. Wagoner, "Simulation of springback," *Int. J. Mech. Sci.*, vol. 44, no. 1, pp. 103–122, Jan. 2002.
- [5] X. A. Yang and F. Ruan, "A die design method for springback compensation based on displacement adjustment," *Int. J. Mech. Sci.*, vol. 53, no. 5, pp. 399–406, May 2011.
- [6] Z. K. Zhang, J. J. Wu, S. Zhang, M. Z. Wang, R. C. Guo, and S. C. Guo, "A new iterative method for springback control based on theory analysis and displacement adjustment," *Int. J. Mech. Sci.*, vol. 105, pp. 330–339, Jan. 2016.
- [7] R. Lingbeek, J. Huétink, S. Ohnimus, M. Petzoldt, and J. Weiher, "The development of a finite elements based springback compensation tool for sheet metal products," *J. Mater. Process. Technol.*, vol. 169, no. 1, pp. 115–125, Oct. 2005.
- [8] J. Lin, T. A. Dean, R. P. Garrett, and A. D. Foster, *A Process in Forming High Strength and Complex-Shaped Al-Alloy Sheet Components*, WO2008059242A2, 2008.
- [9] H. Wang, J. Zhou, T. S. Zhao, and Y. P. Tao, "Springback compensation of automotive panel based on three-dimensional scanning and reverse engineering," *Int. J. Adv. Manuf. Technol.*, vol. 85, no. 5-8, pp. 1187–1193, Jul. 2016.
- [10] Y. Li, Q. Rong, Z. S. Shi, X. G. Sun, L. C. Meng, and J. G. Lin, "An accelerated springback compensation method for creep age forming," *Int. J. Adv. Manuf. Technol.*, vol. 102, no. 1–4, pp. 121–134, May 2019.
- [11] J. Cao, E. Brinksmeier, M. W. Fu, R. X. Gao, B. Liang, M. Merklein, M. Schmidt, and J. Yanagimoto, "Manufacturing of advanced smart tooling for metal forming," *CIRP Ann.*, vol. 68, no. 2, pp. 605–628, Jun. 2019.
- [12] M. Ghahramani, Y. Qiao, M. C. Zhou, A. O'Hagan, and J. Sweeney, "AI-based modeling and data-driven evaluation for smart manufacturing processes," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 4, pp. 1026–1037, Jul. 2020.
- [13] C. Ieracitano, A. Paviglianiti, M. Campolo, A. Hussain, E. Pasero, and F. C. Morabito, "A novel automatic classification system based on hybrid unsupervised and supervised machine learning for electrospun nanofibers," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 1, pp. 64–76, Jan. 2021.
- [14] C. Jaremenko, N. Ravikumar, E. Affronti, M. Merklein, and A. Maier, "Determination of forming limits in sheet metal forming using deep learning," *Materials*, vol. 12, no. 7, pp. 1051, Mar. 2019.
- [15] S. M. Liu, Z. S. Shi, J. G. Lin, and Z. Q. Li, "Reinforcement learning in free-form stamping of sheet-metals," *Procedia Manuf.*, vol. 50, pp. 444–449, Jan. 2020.
- [16] J. W. Xiang, T. Matsumoto, Y. X. Wang, and Z. S. Jiang, "Detect damages in conical shells using curvature mode shape and wavelet finite element method," *Int. J. Mech. Sci.*, vol. 66, pp. 83–93, Jan. 2013.
- [17] M. A. Dib, N. J. Oliveira, A. E. Marques, M. C. Oliveira, J. V. Fernandes, B. M. Ribeiro, and P. A. Prates, "Single and ensemble classifiers for defect prediction in sheet metal forming under variability," *Neural Comput. Appl.*, vol. 32, no. 16, pp. 12335–12349, Aug. 2020.
- [18] X. Y. Liu, Y. J. Du, X. L. Lu, and S. P. Zhao, "Springback prediction and forming accuracy control of micro w-bending using support vector machine," in *Proc. 6th Int. Conf. Frontiers of Industrial Engineering*, London, United Kingdom, 2019, pp. 23–127.
- [19] A. Ben Abdesslem and A. El-Hami, "A probabilistic approach for optimising hydroformed structures using local surrogate models to control failures," *Int. J. Mech. Sci.*, vol. 96-97, pp. 143–162, Jun. 2015.
- [20] M. S. Ashhab, T. Breitsprecher, and S. Wartzack, "Neural network based modeling and optimization of deep drawing – extrusion combined process," *J. Intell. Manuf.*, vol. 25, no. 1, pp. 77–84, Feb. 2014.
- [21] R. Narayanasamy and P. Padmanabhan, "Comparison of regression and artificial neural network model for the prediction of springback during air bending process of interstitial free steel sheet," *J. Intell. Manuf.*, vol. 23, no. 3, pp. 357–364, Jun. 2012.
- [22] Z. F. Guo and W. C. Tang, "Bending angle prediction model based on BPNN-spline in air bending springback process," *Math. Probl. Eng.*, vol. 2017, pp. Article ID 7834621, Feb. 2017.
- [23] V. Viswanathan, B. Kinsey, and J. Cao, "Experimental implementation of neural network springback control for sheet metal forming," *J. Eng. Mater. Technol.*, vol. 125, no. 2, pp. 141–147, Apr. 2003.
- [24] N. Mekras, "Using artificial neural networks to model aluminium based sheet forming processes and tools details," *J. Phys.: Conf. Ser.*, vol. 896, no. 1, pp. 012090, Jul. 2017.
- [25] A. Jenab, I. Sari Sarraf, D. E. Green, T. Rahmaan, and M. J. Worswick, "The use of genetic algorithm and neural network to predict rate-dependent tensile flow behaviour of AA5182-O sheets," *Mater. Des.*, vol. 94, pp. 262–273, Mar. 2016.
- [26] X. Y. Li, C. C. Roth, and D. Mohr, "Machine-learning based temperature- and rate-dependent plasticity model: Application to analysis of fracture experiments on DP steel," *Int. J. Plast.*, vol. 118, pp. 320–344, Jul. 2019.
- [27] C. Hartmann, D. Opritescu, and W. Volk, "An artificial neural network approach for tool path generation in incremental sheet metal free-forming," *J. Intell. Manuf.*, vol. 30, no. 2, pp. 757–770, Feb. 2019.
- [28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017.
- [29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
- [30] Y. F. Xia, H. Yu, and F. Y. Wang, "Accurate and robust eye center localization via fully convolutional networks," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 5, pp. 1127–1138, Sep. 2019.
- [31] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proc.*

- 25th Int. Conf. Machine Learning, 2008, pp. 160–167.
- [32] M. L. Seltzer and J. Droppo, “Multi-task learning in deep neural networks for improved phoneme recognition,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, 2013, pp. 6965–6969.
- [33] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, Massachusetts: The MIT Press, 2016.
- [34] E. Hamouche and E. G. Loukaides, “Classification and selection of sheet forming processes with machine learning,” *Int. J. Comput. Integr. Manuf.*, vol. 31, no. 9, pp. 921–932, Jan. 2018.
- [35] C. Sauer, B. Schleich, and S. Wartzack, “Deep learning in sheet-bulk metal forming part design,” in *Proc. 15th Int. Design Conf.*, Dubrovnik, Croatia, 2018, pp. 2999–3010.
- [36] J. Pfrommer, C. Zimmerling, J. Z. Liu, L. Kärger, F. Henning, and J. Beyerer, “Optimisation of manufacturing process parameters using deep neural networks as surrogate models,” *Procedia CIRP*, vol. 72, pp. 426–431, Mar. 2018.
- [37] C. Zimmerling, D. Trippe, B. Fengler, and L. Kärger, “An approach for rapid prediction of textile draping results for variable composite component geometries using deep neural networks,” *AIP Conf. Proc.*, vol. 2113, pp. 020007, Jul. 2019.
- [38] A. M. Tartakovsky, C. O. Marrero, P. Perdikaris, G. D. Tartakovsky, and D. Barajas-Solano, “Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems,” *Water Resour. Res.*, vol. 56, no. 5, pp. e2019WR026731, May 2020.
- [39] R. Y. Zhang, Y. Liu, and H. Sun, “Physics-informed multi-LSTM networks for metamodeling of nonlinear structures,” *Comput. Methods Appl. Mech. Eng.*, vol. 369, pp. 113226, Sep. 2020.
- [40] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019.
- [41] M. A. Nabian and H. Meidani, “Physics-driven regularization of deep neural networks for enhanced engineering design and analysis,” *J. Comput. Inf. Sci. Eng.*, vol. 20, no. 1, pp. 011006, Feb. 2020.
- [42] H. P. Yao, Y. Gao, and Y. M. Liu, “FEA-Net: A physics-guided data-driven model for efficient mechanical response prediction,” *Comput. Methods Appl. Mech. Eng.*, vol. 363, pp. 112892, May 2020.
- [43] Q. Rong, Y. Li, Z. S. Shi, L. C. Meng, X. H. Sun, X. G. Sun, and J. G. Lin, “Experimental investigations of stress-relaxation ageing behaviour of AA6082,” *Mater. Sci. Eng. A*, vol. 750, pp. 108–116, Mar. 2019.
- [44] H. L. Sun, X. K. Song, R. Ma, and J. Zhao, “Theoretical analysis of JCO forming and springback for sheet metal four-point bending in manufacture of LSAW pipe,” *China Mech. Eng.*, vol. 25, no. 2, pp. 257–262, Mar. 2014.
- [45] Z. Marciniak, J. L. Duncan, and S. J. Hu, *Mechanics of Sheet Metal Forming*. 2nd ed. Oxford: Elsevier, 2002.
- [46] J. Johnson, A. Karpathy, and F. F. Li, “DenseCap: Fully convolutional localization networks for dense captioning,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 4565–4574.
- [47] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
- [48] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” in *Proc. 3rd Int. Conf. Learning Representations*, San Diego, USA, 2015, pp. 1–15. [Online]. Available: <https://arxiv.org/abs/1412.6980>. Accessed: Jan. 15, 2021.
- [49] G. Hinton, N. Srivastava, and K. Swersky, “Neural networks for machine learning: lecture 6a: Overview of mini-batch gradient descent.” [Online]. Available: https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf. Accessed: Feb. 1, 2020.
- [50] R. Caruana, “Multitask connectionist learning,” in *Proc. 1993 Connectionist Models Summer School*, Hillsdale, NJ, USA, 1993, pp. 372–379.
- [51] R. Ranjan, V. M. Patel, and R. Chellappa, “HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 121–135, Jan. 2019.
- [52] Z. X. Wang, X. F. Zhu, E. Adeli, Y. Y. Zhu, F. P. Nie, B. Munsell, and G. R. Wu, “Multi-modal classification of neurodegenerative disease by progressive graph-based transductive learning,” *Med. Image Anal.*, vol. 39, pp. 218–230, Jul. 2017.
- [53] T. Altan and E. Tekkaya, *Sheet Metal Forming: Processes and Applications*. Metals Park, OH: ASM Int., 2012.
- [54] X. Luo, M. C. Zhou, Y. N. Xia, Q. S. Zhu, A. C. Ammari, and A. Alabdulwahab, “Generating highly accurate predictions for missing QoS data via aggregating nonnegative latent factor models,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 3, pp. 524–537, Mar. 2016.
- [55] D. Wu, X. Luo, M. S. Shang, Y. He, G. Y. Wang, and X. D. Wu, “A data-characteristic-aware latent factor model for web services QoS prediction,” *IEEE Trans. Knowl. Data Eng.*, DOI: 10.1109/TKDE.2020.3014302, 2020.
- [56] X. Luo, Y. Yuan, S. L. Chen, N. Y. Zeng, and Z. D. Wang, “Position-transitional particle swarm optimization-incorporated latent factor analysis,” *IEEE Trans. Knowl. Data Eng.*, DOI: 10.1109/TKDE.2020.3033324, 2020.
- [57] X. Luo, H. Wu, H. Q. Yuan, and M. C. Zhou, “Temporal pattern-aware QoS prediction via biased non-negative latent factorization of tensors,” *IEEE Trans. Cybern.*, vol. 50, no. 5, pp. 1798–1809, May 2020.
- [58] B. Shen, Z. D. Wang, and H. Qiao, “Event-triggered state estimation for discrete-time multidelayed neural networks with stochastic parameters and incomplete measurements,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1152–1163, May 2017.
- [59] Y. F. Liu, B. Shen, and H. S. Shu, “Finite-time resilient H_∞ state estimation for discrete-time delayed neural networks under dynamic event-triggered mechanism,” *Neural Netw.*, vol. 121, pp. 356–365, Jan. 2020.



Shiming Liu received the M.Sc. degree from the Department of Mechanical Engineering at Imperial College London, UK, in 2018. He is currently pursuing the Ph.D. degree at Mechanics of Materials Division, Department of Mechanical Engineering, Imperial College London, UK. His research interests include intelligent metal forming system, especially deep learning in sheet metal forming.



Yifan Xia received the M.Sc. degree from Ocean University of China in 2017. He is currently pursuing the Ph.D. degree at the School of Creative Technologies, the University of Portsmouth, UK. His research interests include computer vision, especially eye gaze estimation and facial expression recognition.



Zhusheng Shi is an Advanced Research Fellow in the Department of Mechanical Engineering at Imperial College London, UK. He received the BSc degree from Northeastern University, China, the MSc degree from Central Iron and Steel Research Institute, China, and the Ph.D. degree in metallurgy and materials from the University of Birmingham, UK. His research interests include metal forming, materials characterization, processing and modelling. He has published over 100 research papers.



Hui Yu (SM' 15) is a Professor with the University of Portsmouth, UK. His research interests include methods and practical development in vision, machine learning and AI with applications to human-machine interaction, virtual and augmented reality, robotics and geometric processing of facial expression. He serves as an Associate Editor of *IEEE Transactions on Human-Machine Systems* and *IEEE/CAA Journal of Automatica Sinica*.



Zhiqiang Li is a Professor with Manufacturing Technology Institute (MTI), AVIC. He graduated from Northwestern Polytechnical University in 1986 and then joined MTI, AVIC in the same year. He received the Ph.D. degree from the University of Science and Technology, Beijing. Professor Li has over 30 years solid research experience in manufacturing technology, especially in superplasticity, for aircraft components and equipment. He has rich experience in materials processing and

aeronautical & astronautic manufacturing engineering. His research interests include superplastic forming (SPF)/diffusion bonding (DB) technology of Titanium, forming technologies of complex structures, spinning process and equipment, metal matrix composites fabrication, forming and reinforcement with shot peening technology, stretch forming with reconfigurable tooling.

He is the Visiting Professor at Imperial College London and the Research Committee Member of the Welding Institute (TWI), UK. He established and has been the Editor-in-Chief for the *International Journal of Lightweight Materials and Manufacture*.



Jianguo Lin is a Professor and Royal Academy of Engineering Research Chair in the Department of Mechanical Engineering, Imperial College London, UK. He received the Ph.D. degree from the Department of Mechanical Engineering, University of Sheffield, UK, in 1991. His research interests include metal forming, materials and process modelling, solid/computational mechanics. He has published over 300 research papers.

Since 2008, he is a Chair in the Department of Mechanical Engineering, Imperial College London, UK. In 2008, he established the Metal-Forming and Materials Modelling Group at Imperial College London. He is a Founder Director of 3 research centers and 1 joint research laboratory: AVIC Centre for Structural Design and Manufacture, BIAM – Imperial Centre for Materials Characterization, Processing and Modelling, and CRRC Sifang – Imperial Research Centre for Transportation Manufacturing Technologies. He is also a Founder Director of Impression Technologies Ltd and CurvEx Technology Ltd.

Professor Lin was elected as Fellow of the Royal Academy of Engineering (FREng) in 2013. He is a Member of European Academy of Science, Fellow of IMechE and IMMM, and Vice Chairman of International Group on Modelling Material Flow and Structure Evolution in the Metallurgical Process (MSEMP).